Notes 5a: One-sample t test

1. Purpose

One-sample t-test is designed to test whether one sample of data differs from a standard value or a population mean. The data must be quantitative (ratio, interval, or sometimes ordinal).

Example 1: Do students in EDUR 8131 have a systolic blood pressure that differs from the national average of 120?

Example 2: Does a single-serving sized bag of peanut M&Ms weigh the claimed weight of 1.74oz?

2. Steps of Hypothesis Testing

Like with the one-sample Z test, the one-sample t test follows the same steps for hypothesis testing (and this is true for all hypothesis testing procedures):

   a.   Define both H_0 and H_1
   b.   Set alpha (α, probability of a Type 1 Error)
   c.   Calculate df (degrees of freedom; this step is not part of the one-sample Z test)
   d.   Identify decision rule (either for α, test statistic, or confidence interval)
   e.   Calculate the test statistic (t ratio for one-sample t test)
   f.   Present statistical results and interpretation of results

3. Formula

Before reviewing hypotheses for the one-sample t test, first consider the distinction between the formulas for the one-sample Z and t tests. Recall formula for one-sample Z test:

$$Z_{\bar{x}} = \frac{M - \mu}{\sigma/\sqrt{n}}$$

Formula for one-sample t test is almost identical:

$$t = \frac{M - \mu}{s/\sqrt{n}}$$

Notice the only difference is use of s (the sample standard deviation) instead of σ (population standard deviation). The denominator $s/\sqrt{n}$ is known as the standard error of the mean.

4. Hypotheses

(a) Example 1

Do students in EDUR 8131 have a systolic blood pressure that differs from the national average of 120?

Null
      Written
            The systolic blood pressure for EDUR 8131 students equals the national average of 120.

      Symbolic
            $H_0: \mu = 120$ or $H_0: \mu - 120 = 0$

Alternative Non-directional
      Written
            The systolic blood pressure for EDUR 8131 students differs from the national average of 120.

      Symbolic
            $H_1: \mu \neq 120$ or $H_1: \mu - 120 \neq 0$

(b) Example 2

Does a single-serving sized bag of peanut M&Ms weigh the claimed weight of 1.74oz?

Null
      Written
            Contents of a single-serving sized bag of peanut M&Ms will weigh 1.74 oz.

      Symbolic
            $H_0: \mu = 1.74$ or $H_0: \mu - 1.74 = 0$

Alternative Non-directional
      Written
            Contents of a single-serving sized bag of peanut M&Ms will not weigh 1.74 oz.

      Symbolic
            $H_1: \mu \neq 1.74$ or $H_1: \mu - 1.74 \neq 0$

5. Alpha (α), Degrees of Freedom, and Critical Values

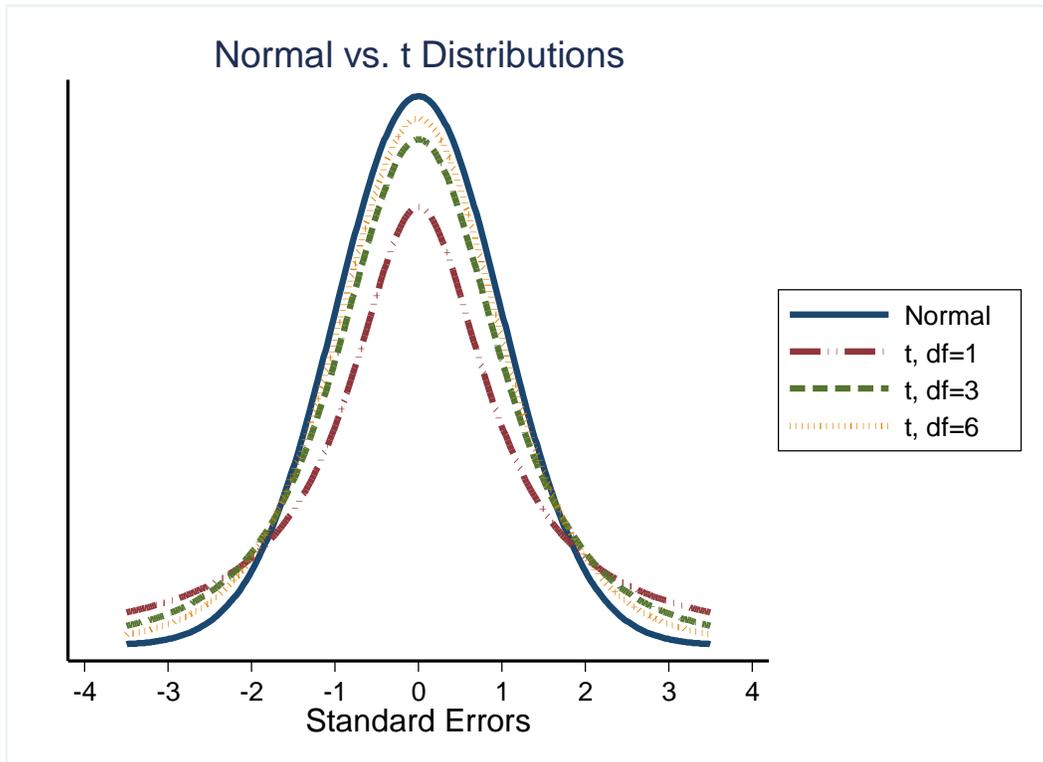Recall that with a one-sample Z test the critical values were

      $\alpha = .10$, then critical $Z \approx \pm 1.65$
      $\alpha = .05$, then critical $Z = \pm 1.96$
      $\alpha = .01$, then critical $Z \approx \pm 2.58$

Where, as noted above, α is the probability of making a Type 1 Error (claiming an effect or relationship exists in the population when it does not). Normally this probability is set low to .01 or .05.

One cannot use the normal distribution Z table to find critical t-values because the t distribution, while approximately normal, has thicker tails. The graph below shows examples of three t-distributions compared with a standard normal distribution. Note the tails have slightly more area therefore the critical values will be larger, in absolute value, for a given tail area alpha value.

Normal vs. t Distributions

Legend:
- Normal
- t, df=1
- t, df=3
- t, df=6

Standard Errors (x-axis: -4, -3, -2, -1, 0, 1, 2, 3, 4)

To find critical t-values one must first calculate degrees of freedom

df (or sometimes Greek nu, ν) = n − 1

and then locate critical t values via use of t table or relevant software.

Example 1: Eight students sampled to test systolic BP. What is the appropriate critical t value for each Type 1 error rate?

   df = 8 − 1 = 7

   α = .10, then critical t = $\pm$ 1.89
   α = .05, then critical t = $\pm$ 2.36
   α = .01, then critical t = $\pm$ 3.50

Example 2: Ten peanut M&M bags purchased and contents weighed. What is the appropriate critical t value for each Type 1 error rate?

   df = 10 − 1 = 9

   α = .10, then critical t = $\pm$ 1.83
   α = .05, then critical t = $\pm$ 2.26
   α = .01, then critical t = $\pm$ 3.25

Exercise—find critical t values for the following:

1. n = 52, α = .05   df = 52 − 1 = 51, critical t = ±2.01
2. n = 120, α = .01   df = 120-1 = 119, critical t = ±2.63
3. n = 158, α = .05   df = 158-1 = 157, critical t = ±1.98

6. Decision Rules

Like with the one-sample Z test, there are decision rules for determining whether to reject the null hypothesis.

(a) P-values and α

Decision rule for p-values (this decision rule holds for all statistical tests in which p-values are present):

> *If p ≤ α reject H$_0$, otherwise fail to reject H$_0$*

A p-value is calculated for the sampled data assuming the null hypothesis is true. A p-value for the one-sample t test may be interpreted as follows:

> *p-value*: The probability of obtaining a sample mean that deviates, in a random fashion, from the population value (μ) by the amount observed, or to a greater extent, assuming the null hypothesis is true (i.e., H$_0$: μ = comparison value).

Thus the p-value indicates the probability of obtaining random sample means that will deviate randomly (in either direction) this far or further from the population value if, in fact, there should be no difference between population μ and the value specified in H$_0$.

(b) Critical t values (t$_{crit}$) and calculated t

An equivalent hypothesis approach is to compare calculated t-values against critical t-values:

> *If -t ≤ -t$_{crit}$ or t ≥ t$_{crit}$ reject H$_0$, otherwise fail to reject H$_0$*

This can be simplified by using absolute values as follows:

$$If \ |t| \geq |t_{crit}| \ reject \ H_0, otherwise \ fail \ to \ reject \ H_0$$

(c) Confidence Intervals

This will be explained below in section 9.

7. Example t-tests (with SPSS)

(a) Example 1

Do students in EDUR 8131 have a systolic blood pressure that differs from the national average of 120?

Systolic Data for eight students

     139 145 153 115 175 146 137 161

Summary Statistics: M = 146.375, SD = 17.752

$$t = \frac{M - \mu}{s/\sqrt{n}} = \frac{146.375 - 120}{17.752/\sqrt{8}} = \frac{26.375}{6.276} = 4.202$$

If α = .05 and df = 8 − 1 = 7, then the critical t = ± 2.36. Therefore, using the absolute value decision rule for t-values:

    If 4.202 ≥ 2.36 reject H$_0$, otherwise fail to reject H$_0$

Since 4.202 is larger than 2.36 the null is rejected.

> Work this example in SPSS to show how one obtains p-values with one sample t-test.

SPSS provides p-value of .004, so:

    If .004 ≤ .05 reject H$_0$, otherwise fail to reject H$_0$

Since .004 is less than .05 the null is rejected.

Reject H$_0$ and conclude that students in EDUR 8131 have a systolic blood pressure that is higher than the national mean of 120.

(b) Example 2

Do the contents of a single-serving sized bag of peanut M&Ms weigh the claimed weight of 1.74oz?

Content weight of ten single-sized serving bags of peanut M&Ms (in oz)

    1.81 1.69 1.71 1.55 1.72 1.76 1.79 1.68 1.70 1.68

Summary Statistics: M = 1.709, SD = 0.072

$$t = \frac{M - \mu}{s/\sqrt{n}} = \frac{1.709 - 1.74}{0.072/\sqrt{10}} = \frac{-0.031}{0.0228} = -1.359$$

If α = .01 and df = 10 − 1 = 9, then critical t = ± 3.25. Therefore, using the absolute value decision rule for t-values:

    If 1.359 ≥ 3.25 reject H$_0$, otherwise fail to reject H$_0$

Since 1.359 is not greater than 3.25 fail to reject the null.

SPSS provides p-value of .207, so:

> *If .207 ≤ .01 reject H₀, otherwise fail to reject H₀*

Since .207 is not less than .01 fail to reject the null.

Fail to reject H_0 and conclude that content weight of single-serving bags of peanut M&Ms appears to be consistent with the claimed weight of 1.74oz.

(c) Example 3

Fifteen students have a sample MAT mean of 32.3 with a sample standard deviation of 4.73. The national mean MAT is 30.

(c.1) Does this sample appear to have a mean that differs from the national mean at the .01 level of significance?

(c.2) What about at the .10 level of significance?

Answer (from Stata software):

```
. ttesti 15 32.3 4.73 30

  Variable |       Obs        Mean     Std. Dev.
  ---------+------------------------------------
         x |        15        32.3         4.73

        Ho:  mean = 30
                  t = 1.88 with 14 d.f.
           Pr > |t|  = 0.0806
```

8. Confidence Intervals

(a) Confidence Interval about the Sample Mean

The formula for confidence intervals about a sample mean is

$$(1-\alpha)CI = \bar{X} \pm (t_{df})(s_{\bar{x}})$$

Using words, this formula becomes

$$(1-\alpha)CI = \bar{X} \pm (critical\ t\ value)(standard\ error\ of\ mean)$$

and therefore reduces to three components: sample mean, critical t value, and standard error of the mean (the denominator in the t-ratio formula).

Recall that the standard error of the mean is just the sample SD divided by the square root of the sample size

$$s_{\bar{x}} = \frac{s}{\sqrt{n}}$$

Example 1: 95% Confidence interval for blood pressure mean

Statistics for blood pressure:

M = 146.375
SD = 17.752
N = 8
α = .05
critical t = $\pm$ 2.36

The standard error of the mean is

$$s_{\bar{x}} = \frac{s}{\sqrt{n}} = \frac{17.752}{\sqrt{8}} = 6.276$$

The 95% CI formula

$$(1-\alpha)CI = \bar{X} \pm (critical\ t\ value)(standard\ error\ of\ mean)$$

$$(1-.05)CI = 146.375 \pm (2.36)(6.276)$$

$$.95\ CI = 146.375 \pm (2.36)(6.276)$$

Upper Limit = $146.375 + (2.36)(6.276) = 146.375 + 14.811 = 161.186$
Lower Limit = $146.375 - (2.36)(6.276) = 146.375 - 14.811 = 131.564$

Interpretation
> One may be 95% confident that the interval 131.564 to 161.186 includes the population mean. (Note that these upper and lower limits are not consistent with a value of 120 – the interval does not contain the hypothesized value of 120; this inconsistency will be discussed below in the form of hypothesis testing.)

Example 2: 99% Confidence interval for content weight of single-serving bag of peanut M&Ms

Statistics for M&M content weight:

M = 1.709
SD = 0.072
N = 10
α = .01
critical t = $\pm$ 3.25

The standard error of the mean is

$$s_{\bar{x}} = \frac{s}{\sqrt{n}} = \frac{0.072}{\sqrt{10}} = 0.0228$$

The 99% CI formula

$$(1 - \alpha)CI = \bar{X} \pm (critical\ t\ value)(standard\ error\ of\ mean)$$

$$(1 - .01)CI = 1.709 \pm (3.25)(0.0228)$$

$$.99\ CI = 1.709 \pm (3.25)(0.0228)$$

Upper Limit = $1.709 + (3.25)(0.0228) = 1.709 + 0.0741 = 1.783$
Lower Limit = $1.709 - (3.25)(0.0228) = 1.709 - 0.0741 = 1.635$

SPSS: To obtain confidence intervals about means, use the Explore command in SPSS.

Note: SPSS calculates the 99% CI as 1.657, 1.761; the observed difference between these values and the values above are due to rounding of the critical t value and the standard error of the mean.

Interpretation
> One may be 99% confident that the population mean for content weight for a single-serving bag of M&Ms is within the limits of 1.635 and 1.783. (Note that this range is consistent with the claim of 1.74—the value of 1.74 is within the interval calculated.)

(b) Confidence Interval about the Mean Difference (SPSS—one-sample t test style)

Example 1: 95% Confidence interval for blood pressure mean difference $(\bar{X} - \mu)$

For the blood pressure data the

95% CI = 131.564, 161.186

However, when running the one-sample t test, SPSS reports the 95% CI as

95% CI = 11.53, 41.22

Why the difference?

Recall that the null hypothesis may be symbolized one of two ways

H₀: μ = 120   or   H₀: μ − 120 = 0

The 95% CI we calculated is based upon H₀: μ = 120, and the 95% CI SPSS presents is based upon H₀: μ − 120 = 0, so SPSS is calculating the CI for the mean difference $(\bar{X} - \mu)$.

The CI formula for mean difference:

$$(1 - \alpha)CI = (\bar{X} - \mu) \pm (critical\ t\ value)(standard\ error\ of\ mean)$$

where μ is the test value under consideration. For example, using the blood pressure data, the 95%CI on the mean difference is

$$(1 - \alpha)CI = (\bar{X} - \mu) \pm (critical\ t\ value)(standard\ error\ of\ mean)$$

$$(1 - .05)CI = (146.375 - 120) \pm (2.36)(6.276)$$

$$.95\ CI = \ 26.375 \pm 14.811$$

Upper Limit = 26.375 + (2.36)(6.276) = 26.375 + 14.811 = 41.186
Lower Limit = 26.375 − (2.36)(6.276) = 26.375 − 14.811 = 11.564

These values, 11.56 to 41.18, agree with SPSS's 95% CI for the mean difference within rounding error.

Interpretation
> One may be 95% confident that systolic blood pressure population mean difference from a value of 120 is within the interval of 11.56 to 41.18.

Example 2: 99% Confidence interval for content weight of single-serving bag of peanut M&Ms

Recall the null hypothesis for the M&M data: H₀: μ = 1.74   or   ***H₀: μ − 1.74 = 0***

Statistics for M&M content weight:

M = 1.709
μ = 1.74
SD = 0.072
N = 10
α = .01
critical t = ± 3.25

The standard error of the mean is

$$s_{\bar{x}} = \frac{s}{\sqrt{n}} = \frac{0.072}{\sqrt{10}} = 0.0228$$

The 99% CI formula

$$(1 - \alpha)CI = (\bar{X} - \mu) \pm (critical\ t\ value)(standard\ error\ of\ mean)$$
$$(1 - .01)CI = (1.709 - 1.74) \pm (3.25)(0.0228)$$
$$.99\ CI = -0.031 \pm 0.07399$$

Upper Limit = $-0.031 + (3.25)(0.0228) = -0.031 + 0.07399 = 0.042997$
Lower Limit = $-0.031 - (3.25)(0.0228) = -0.031 - 0.07399 = $ -0.10499

Interpretation

> One may be 99% confident that M&M weight difference between population value of 1.74oz and the population mean weight from which our sample was taken could be between -0.105oz to 0.043oz.

9. Hypothesis Testing with CI

In addition to use of critical t values and p-values, one may perform hypothesis testing with confidence intervals. The logic depends upon whether the CI was calculated for a mean or a mean difference. Both are explained below.

Key Concept: Recall that the confidence interval presents a range of possible values for the population parameter – these are values we believe are likely estimates for the parameter. If the value hypothesized in the null (e.g., H_0: μ = 120 or H_0: μ - 120 = 0) is not within the range specified by the confidence interval, then that hypothesized value does not represent one of the likely values for the population parameter.

(a) Confidence Interval about the Sample Mean

Recall the hypothesized value of the systolic blood pressure mean:

H_0: μ = 120

This hypothesis states that the population from which our sample was drawn will have a systolic mean will equal 120.

If the confidence interval contains this value, then one may conclude that the sampled data is consistent with a population value blood pressure of 120 (i.e., fail to reject the null).

If the confidence interval does not contain the value 120, then one may conclude the sample data do not come from a population with a mean value of 120 (i.e., reject the null).

95% CI for blood pressure: 131.56 to 161.18
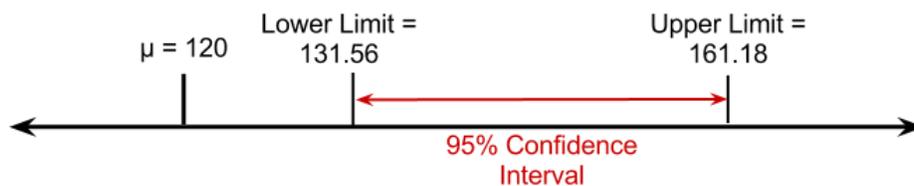
The figure above shows that the value of 120 is not within the 95% confidence interval.

Since the value 120 does not appear within the interval 131.56 to 161.18 we may conclude that the sampled data come from a population with a mean systolic blood pressure that is higher than 120, so reject the null. That is, the value of 120 does not appear to be one of the values for which we are 95% confident for the population mean blood pressure for EDUR 8131 students.

The logic: the confidence interval provides a range of values in which we can be confident the population parameter will appear, so if the hypothesized value stated by the null (e.g., H_0: μ = 120) is not found within the confidence interval, then that value (e.g., 120) is not one in which we can be confident for the population sampled thus we can reject that value.

(b) Confidence Interval about the Mean Difference (SPSS—one-sample t test style)

Example 1: Blood Pressure

For the blood pressure data an alternative null specifies the mean systolic blood pressure does not differ from 120:

$H_0: \mu - 120 = 0$

Since this hypothesis specifies the mean difference to be 0.00, then if the CI for the mean difference includes the value of 0.00 one would fail to reject the null hypothesis $H_0: \mu - 120 = 0$ (i.e., conclude the sampled data are consistent with a mean difference of 0.00).

But if the CI does not include the value 0.00, one would reject Ho and conclude the sampled data have a mean difference that is either larger or smaller than 0 (i.e., sampled data mean difference is not consistent with a value of 0.00).

The 95% CI for mean difference in blood pressure is 11.56 to 41.18.
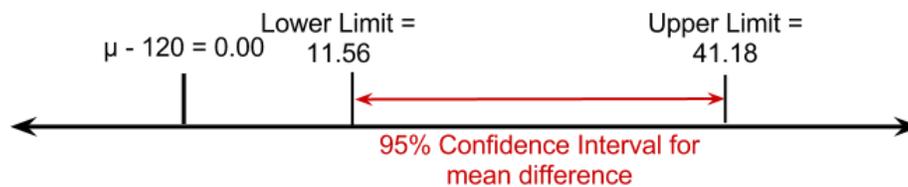
As the above figure shows, the value of 0.00 does not lie within the 95% CI for the mean difference.

Since 0.00 is not included in the range 11.56 to 41.18 we conclude that a value of 0.00 for the mean difference is not likely – is not reasonable – so we reject Ho and conclude the mean difference appears to be larger than 0.00 since the range 11.56 to 41.18 is larger than 0.00.

Example 2: M&M Weight Data

For M&M data the null states: $H_0: \mu - 1.74 = 0.00$

The 99% CI for mean difference ranges from -0.105 to 0.043.

Since 0.00 lies within this interval, as shown by the figure below, one would conclude that 0.00 is one of the possible values for the population mean difference therefore 0.00 cannot be rejected as unlikely, so $H_0: \mu - 1.74 = 0.00$ would not be rejected.
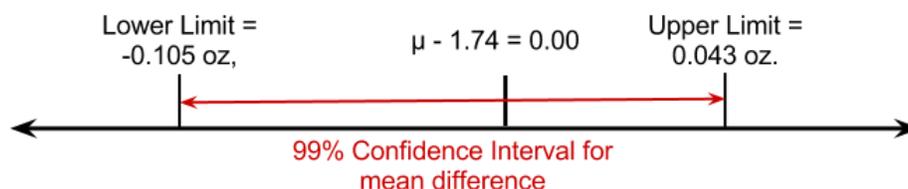
In short, if one
- calculates a confidence interval for the mean difference,
- finds that 0.00 lies within that interval then one fails to reject the null, but if
- finds that 0.00 lies outside that interval then one rejects the null

10. Assumptions of One-sample t-test

One-sample t-test has the same assumptions as the one-sample Z test:

- data were randomly sampled,
- from a population of scores that are normally distributed,
- data are independent, and
- data are suitable for calculation of mean (i.e., data are interval or ratio scale of measurement, but ordinal data often works well too).

If the data were randomly sampled, then independence can be assumed, and if the sample is large enough (~25 or more), then the t-test is robust to the assumption that the population scores are normally distribution due to the central limit theorem.

11. APA Style Results Presentation

It is possible to report one-sample t-test results in either text or tabular format, but for this course present results in tabular format. Both options are provided below.

(a) Text Format

With either format, text or table, one should present written results that include an inferential component (i.e., explain whether the null was rejected) and interpretational component (i.e., explain what results mean). The inferential component will normally have some wording about statistical significance or non-significant results. Recall that significant means only that the null was rejected; it does not mean something important was found. This wording tells readers the null was rejected with an Type 1 error rate of .05: "There was a statistically significant mean difference, at the .05 level, in response rates." This sentence indicates the null was not rejected: "There was no statistically significant mean difference, at the .05 level, in response rates." The interpretation component will inform readers what the results mean in simple language. For example: "Response rates for students in class A were higher than for students in class B."

Below is a sample results presentation for the systolic blood pressure data.

Systolic blood pressure for the sample of EDUR 8131 students was statistically different, at the .05 of significance, from the normed value of 120 (M = 146.38, SD = 17.75, n = 8, 95%CI on mean difference = 11.53 to 41.22, t = 4.20, df = 7, p < .05). Results show that EDUR 8131 students have a higher mean systolic blood pressure than people nationwide.

(b) Tabular Format

Sometimes it is more efficient and easier to comprehend if results are presented in tabular format with corresponding text presentation. Below is an example for both the blood pressure and M&M data.

Table 1
Results of One-sample t-test and Descriptive Statistics for Systolic Blood Pressure and M&M Bag Content Weight

Outcome	M	SD	n	Comparison Value	95% CI for Mean Difference	t	df
Systolic Blood Pressure	146.38	17.75	8	120	11.53, 41.22	4.20*	7
M&M Bag Content Weight	1.71	0.07	10	1.74	-0.08, 0.02	-1.39	9

* $p < .05$.

Results for both the systolic blood pressure and M&M content weight studies are reported in Table 1. Systolic blood pressure for the same of EDUR 8131 students was statistically different, at the .05 level of significance, from the normed value of 120. Results show that EDUR 8131 students have a higher mean systolic blood pressure than people nationwide.

For the M&M study, t-test results show that content weight does not differ statistically from the advertised value of 1.74oz. This finding suggests M&M single-serving bags tend to contain about 1.74 oz. of peanut M&Ms, on average.

12. Topics not Covered

(a) Directional tests

Lower-tailed test
H_1: $\mu < 100$
H_0: $\mu \geq 100$ or

Upper-tailed test
H_1: $\mu > 100$
H_0: $\mu \leq 100$ or

These are not covered because I think non-directional tests are better since don't offer the temptation to change the direction of one's test to obtain statistical significance (i.e., adopts in a post-hoc fashion directional tests since they offer more power compared with non-directional tests). Also statistical software tends to provide p-values only for non-directional tests.

(b) Effect Size d

The effect size d represents the mean difference divided by the group standard deviation:

$$d = \frac{\bar{X} - \mu}{s}$$

Like a Z score, this value indicates the size of the mean difference in standard deviation units. The d formula may appear to be the same as the t formula, but note the different denominator. In the d formula the denominator is s, the sample standard deviation, and in the t formula the denominator is $s_{\bar{x}}$, standard error of the mean.

(c) Sample Size

One may calculate required sample size for a one-sample t test to meet specifications for Type 1 error rate (α), power (1-β), and effect size (d). This requires use of tabled values or software.

13. Exercises

(1) Raw MAT scores are 31, 38, 27, 41, 39, and 36. Is this sample statistically different from the national average MAT of 30? Set $\alpha = .01$.

(2) Same scores, but $\alpha = .05$.

(3) Suppose the following random sample of ITBS-math scores are observed in your middle school: 45, 58, 65, 63, 35, 43, 78, 55, 58, 69, 81, and 49. Is this evidence that your middle school has a student population differs from the national average in terms of mathematics skills (national average ITBS-math is 50)? Set alpha at .05.

(4) You wish to determine whether you are getting cheated every time you buy a bag of apples. The standard bag of apples that you buy states that it contains one pound (16 ounces) of apples. After you get home you notice that the bag only contains 15.5 ounces, not the stated 16 ounces. To determine whether or not the company is systematically cheating the consumer, you decide to buy every 16 ounce bag of apples in the three local grocery stores. After weighing each bag you find the following weights: 14.3, 15.5, 16.3, 17.0, 15.2, 15.9, 14.8, 15.0, 15.2, 15.9, 15.7, 15.6, and 16.1. Setting the significance level at .05; does it seem the company is providing bags that systematically differ from the claimed weight of 16 oz.?

(5) Ford claims that its new car, Aspire, gets 39 mpg on the highway. Consumer Reports magazine wishes to test this claim, so they hire you for $1500 to perform the statistical testing. They buy 10 Aspires and road test each. They find the following mpg estimates for the cars: 32, 43, 39, 38, 34, 36, 35, 38, 39, and 36. Their question to you is: Does our sample of Aspires have an estimated mpg that differs from Ford's claim? Set alpha at .05 and give them an answer.

Answers to Exercises

Computer output from Stata for exercises 1 through 5. Note that "Pr > |t|" is Stata's way of symbolizing the calculated p-value for these tests. Use the t-ratios presented to assess the accuracy of the t-ratio you calculate.

Exercises 1 and 2
```
. ttest mat=30
  Variable |       Obs        Mean    Std. Dev.
  ---------+-------------------------------
      mat |        6     35.33333    5.316641

           Ho:  mean = 30
                   t = 2.46 with 5 d.f.
             Pr > |t| = 0.0574
```

Exercise 3
```
. ttest itbs=50
  Variable |       Obs        Mean    Std. Dev.
  ---------+-------------------------------
     ITBS |       12       58.25     13.93573

           Ho:  mean = 50
                   t = 2.05 with 11 d.f.
             Pr > |t| = 0.0649
```

Exercise 4
```
. ttest weight = 16

  Variable |        Obs           Mean    Std. Dev.
 ---------+-------------------------------
    weight |         13      15.57692     .7013722

          Ho:   mean = 16
                  t = -2.17 with 12 d.f.
           Pr > |t|  = 0.0503
```

Exercise 5
```
. ttest mpg = 39

  Variable |        Obs           Mean    Std. Dev.
 ---------+-------------------------------
       mpg |         10             37     3.091206

          Ho:   mean = 39
                  t = -2.05 with 9 d.f.
           Pr > |t|  = 0.0711
```