Section 9.3 - Tests About a Population Mean (pp. 574-601)

Guidelines for the Normal condition from Chapter 8: (t - PROCEDURES)

- *Sample size less than 15:* IF DATA ROUGHLY SYM, UNIMODAL, NO O/L'S. SKEWED OR O/L'S - DO NOT USE.
- *Sample size at least 15:* USE EXCEPT IF STRONG SKEWNESS OR OUTLIERS
- *Large samples:* $n \geq 30$ EVEN IF SKEWED.

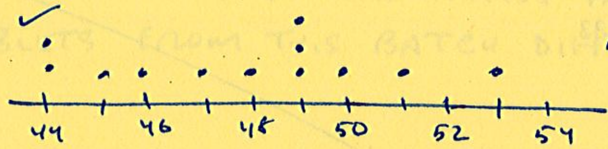
Example - A classic rock (get it?) radio station claims to play an average of 50 minutes of music every hour. However, it seems that every time you turn to this station, there is a commercial playing. To investigate their claim, you randomly select 12 different hours during next week and record what the radio station plays in each of the 12 hours. Here are the number of minutes of music in each of these hours:

44 49 45 51 53 49 44 47 50 46 48 49 $n = 12$

Check the conditions for carrying out a significance test of the company's claim that it plays an average of 50 minutes of music per hour.

RANDOM: SRS ✓

NORMAL:



ROUGHLY SYM. NO OUTLIERS. ⇒ APPROX NORMAL

IND: $10(12) = 120$ MORE THAN 120 HRS / WK.

Test Statistic and *P*-Value - When performing a significance test, we do calculations assuming the null hypothesis is true. The test statistic measures how far the sample result diverges from the parameter value specified by the null hypothesis, in standardized units.

$$\boxed{\text{test statistic} = \frac{\text{statistic} - \text{parameter}}{\text{standard deviation of statistic}}}$$

$$t_{n-1} = \frac{\bar{x} - \mu_0}{s_x/\sqrt{n}}$$

Example (cont) - In our music example, what should the hypotheses be?

$H_0: \mu = 50$

$H_A: \mu < 50$

Compute the test statistic and *P*-value for these data.

$\bar{x} = 47.9 \quad S_x = 2.81$

$t = \dfrac{\bar{x} - \mu_0}{s_x/\sqrt{n}} = \dfrac{47.9 - 50}{2.81/\sqrt{12}} \approx -2.59$

$df = 12 - 1 = 11$

$P(t < -2.59) = 0.0126$

Comments on Table B

USE CALCULATOR!!!

- Table B gives a range of possible *P*-values for a significance test. We can still draw a conclusion from the test in much the same way as if we had a single probability.
- Table B has other limitations.
    - The table does not list all degree of freedom values.
    - The table only shows probabilities for positive values of *t*.

Example

(a) Find the *P*-value for a test of H_0: $\mu = 10$ versus H_A: $\mu > 10$ that uses a sample of size 75 and has a test statistic of $t = 2.33$.

(b) Find the *P*-value for a test of H_0: $\mu = 10$ versus H_A: $\mu \neq 10$ that uses a sample of size 10 and has a test statistic of $t = -0.51$.

Technology - Given the limitations of Table B, it is probably best to use technology to find *P*-values when carrying out a significance test about a population mean.

```
tcdf(1.54,100,14
)
        .0729268628
2*tcdf(-100,-3.1
7,36)
        .0031080065
```

[2nd] [VARS] (DISTR) 5:tcdf(lower bound, upper bound, df)

CYU p. 579

① $H_0: \mu = 320$ vs. $H_A: \mu \neq 320$, where μ is true mean am't of active ingredient (mg) contained in Aspro tablets from this batch of production.

② Random: random sample ✓
Normal: ~~36 < 100~~ 36 ≥ 30 (CLT) ✓
Independent: 36 < 10% of all tablets ✓
} ∴ conditions met.

③ $$t = \frac{319 - 320}{3/\sqrt{36}} = -2$$

④ df = 36 - 1 = 35 (TBL: p-value between 0.05 and 0.10.)
Tech: 2(0.0267) = 0.0534.
Since p-value = 0.0534 > α = 0.05, we fail to reject H_0. There is not convincing evidence that the true mean am't of the active ingredient in Aspro tablets from this batch differs from 320 mg.

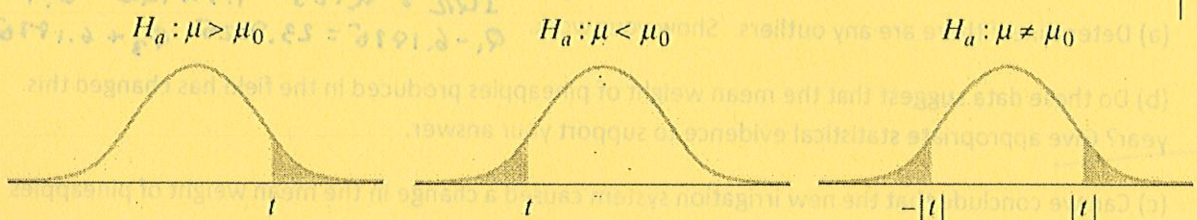
Application - Complete CYU on p. 579.

One-Sample t Test

Choose an SRS of size *n* from a large population with unknown mean μ. To test the hypothesis $H_0: \mu = \mu_0$, compute the one-sample *t* statistic:

$$t_{n-1} = \frac{\bar{x} - \mu_0}{s_x/\sqrt{n}}$$

Find the *P-value* by calculating the probability of getting a *t* statistic this larger or larger in the direction specified by the alternative hypothesis H_A in a *t* distribution with df = *n* - 1.



Use this test only when: (1) the population is Normal or the sample is large (n ≥ 30), and (2) the population is at least 10 times as large as the sample.

Example - Every road has one at some point -- construction zones that have much lower speed limits. To see if drivers obey these lower speed limits, a police officer used a radar gun to measure the speed in miles per hours of a random sample of 10 drivers in a 25 mph construction zone. Here are the results:

27 33 32 21 30 30 29 25 27 34 $n = 10$

(a) Can we conclude that the average speed of drivers in this construction zone is greater than the posted 25 mph speed limit? (b) Given your conclusion in part (a), which kind of mistake -- Type I or Type II -- could you have made? Explain what this mistake means in this context.

A

1. $H_0: \mu = 25$
   $H_A: \mu > 25$
   μ = AVG SPEED OF DRIVERS IN CONST ZONE.
   $\alpha = 0.05$

2. PLAN 1-SAMPLE t-TEST FOR μ:
   RANDOM: ✓
   NORMAL:

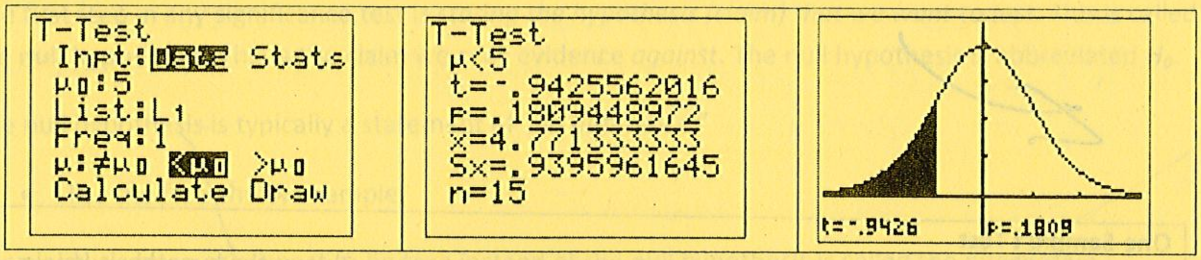
   ROUGH. SYM / NO O/L'S ✓
   IND: MORE THAN 10(10) = 100 DRIVERS ✓

3. DO: $\bar{x} = 28.8$ $S_x = 3.9$ $n = 10$
   TEST STAT: $t = \frac{28.8 - 25}{3.9/\sqrt{10}} = 3.05$
   P-VALUE: $df = 10 - 1 = 9$
   $P(t > 3.05) = 0.0069$.

4. SINCE P-VALUE $= 0.0069 < \alpha = 0.05$ REJECT H_0. CONCLUDE THERE IS CONVINCING EVIDENCE THAT SPEED IS > 25.

B TYPE I SINCE REJECT H_0. WE COULD HAVE CONCLUDED SPEED OF DRIVERS IS > 25 WHEN IT REALLY IS NOT.

Technology



Example (2-sided test) - We will consider the example on pp. 583-584.

Descriptive Statistics: Weight (oz)

Variable	N	Mean	SE Mean	StDev	Minimum	Q1	Median	Q3	Maximum
Weight (oz)	50	31.935	0.339	2.394	26.491	29.990	31.739	34.115	35.547

NO O/L's

(a) Determine if there are any outliers. Show your work.

IQR = 4.125 $\quad 1.5 \times IQR = 6.1875$

$Q_1 - 6.1875 = 23.8025 \quad Q_3 + 6.1875 = 40.3025$

(b) Do these data suggest that the mean weight of pineapples produced in the field has changed this year? Give appropriate statistical evidence to support your answer.

(c) Can we conclude that the new irrigation system caused a change in the mean weight of pineapples produced? Explain your answer.

b) $H_0: \mu = 31$ RANDOM ✓

$H_A: \mu \neq 31$ NORMAL $n > 30$ ✓ (CLT)

$\alpha = 0.05$ IND $10(50) = 500$ PA's ✓

DO: $\bar{x} = 31.935 \quad s_x = 2.394$

TEST STAT: $t = \dfrac{\bar{x} - \mu_0}{s_x/\sqrt{n}} = \dfrac{31.935 - 31}{2.394/\sqrt{50}} = 2.762$

P-VALUE: DF $= 50 - 1 = 49 \quad P(t > 2.762) = 0.0040$

P-VALUE $= 2(.004) = 0.008$.

CONC: SINCE P-VALUE $= .008 < \alpha = 0.05$, REJECT H_0 AND CONCLUDE MEAN WT OF THIS YR'S CROP IS NOT 31 OZ'S.

c) NO, NOT AN EXP!

Confidence Intervals - The connection between two-side tests and confidence intervals is even stronger for means than it was for proportions. This is because both inference methods for means use the standard error of $\bar{x}$ in the calculations:

TEST STAT: $t = \dfrac{\bar{x} - \mu_0}{s_x/\sqrt{n}}$ CI: $\bar{x} \pm t^* \dfrac{s_x}{\sqrt{n}}$

When the two-sided significance test at level α rejects H_0: $\mu = \mu_0$, the 100(1 - α)% confidence interval for μ will not contain the hypothesized value μ_0. And when the test fails to reject the null hypothesis, the confidence interval will contain μ_0.

Tests for Paired Data - We will consider the example on pp. 586-587. (MATCHED PAIRS DESIGN)

(PAIRED t- PROCEDURES)

Using Tests Wisely

- When a null hypothesis can be rejected at the usual levels of significance (0.05 or 0.01), there is good evidence of a difference. But that difference might be very small. When large samples are available, even tiny deviations from the null hypothesis will be significant.
- Statistical significance is not the same thing as practical importance. Pay attention to the data -- do not just focus on the P-value. A few outliers can produce highly significant results.
- The foolish user of statistics who feeds the data to a calculator or computer without exploratory analysis will often be embarrassed.
- Do not ignore lack of significance. When planning a study, verify that the test you plan to use has a high probability (power) of detecting a difference of the size you hope to find.
- Statistical inference is not valid for all sets of data. Badly designed surveys or experiments often produce invalid results. Formal statistical inference cannot correct these flaws.

HW: Read Sec 9.3; problems 59-61,75, 85, 87, 89, 91, 92, 93