Statistical Tests Involving Population Means

Hypothesis test procedures can be applied to any population parameter of interest. In these next few slides we investigate the hypothesis test setting for population means.

Hypothesis Test: Population Means

- Hypothesis tests for population means are carried out using the same 7-step process that was discussed for testing procedures involving proportions.
- The only difference between the two techniques is that there are different “formulas” involved for the point estimator and the test statistic

Statistical Tests Example 1

Back in the late 60’s and early 70’s the sport fishing industry in NY state was concerned about the growing number of lake trout that had elevated levels of mercury (Hg) and cadmium (Cd). In fact the NY department of wildlife issued a strong warning regarding the consumption of lake trout over a certain size. Since that time many changes have taken place and wildlife department personnel would like to test whether the levels of Cd and Hg in lake trout have decreased on average.

Hypothesis Tests for Means

Let’s say that a SRS of 30 lake trout yielded a mean metals concentration of 10.2 ppm with a corresponding standard deviation of 10.8 ppm. Is this enough evidence to conclude that the average heavy metal concentration has fallen since the late 60’s?

Statistical Tests Step 1

- This is the “heavy metal” example. Let’s say that, on average, lake trout contained about 15 ppm heavy metals. Set up the null and alternate hypothesis for testing whether this level has decreased.
- Ho: $\mu \geq 15$
- Ha: $\mu < 15$

Hypothesis Test Process Step 2

- Determine if your test is one tailed or two tailed.
- You can answer this by looking at the alternate hypothesis, Ha. If the inequality points in only one direction then you have a 1 tailed test situation.
- In our example: Ha: $\mu < 15$ hence we have a 1 tailed (left) test.

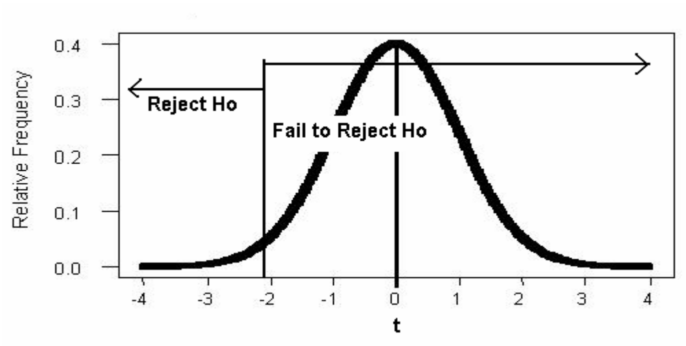
Hypothesis Test Process Step 2A

- Select a significance level for your test.
- The significance level is called **<u>alpha</u> (α).** It is the chance you’re willing to take to arrive at the “wrong” conclusion.
- The level of the test must be specified in advance.
- Let’s use the 0.025 level of significance for this test.

Hypothesis Test Process Step 2B

- If you have a 1 tailed testing situation look to the bottom of the t-table for the column headed by the chosen α level for 1 tail. Here an alpha level of 0.025 was selected.
- Go to the row for 29 df in the column to find t_{crit}
- This value is: 2.05

Step 3: Draw A Decision Graphic



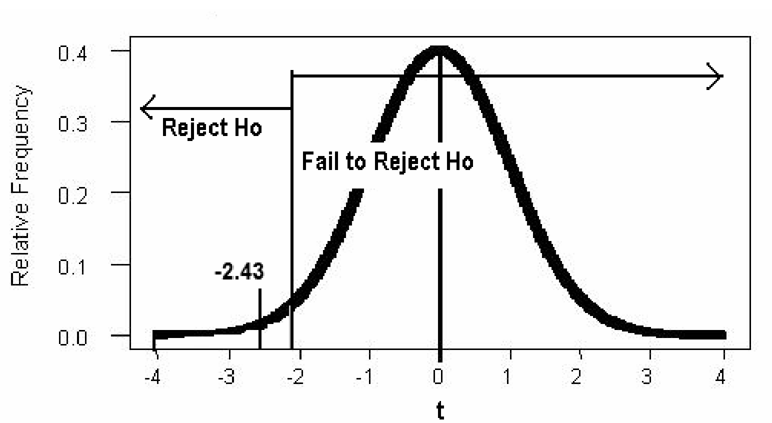
We want to draw a graphic that labels the critical value as well as the rejection region(s)

Step 4: Compute the test statistic

$$t_{calc} = \frac{\bar{x} - \mu_0}{s / \sqrt{n}} = \frac{10.2 - 15}{10.8 / \sqrt{30}} = -2.43$$

We use this formula to compute the test statistic based on our data. After plug-n-chug, t-calc is: -2.43

Step 5: Locate t-calc



Step 6 : The statistical Decision

At this stage our only choice is to

- Reject Ho or
- Fail to reject Ho

- Since t-calc falls in the rejection region we reject Ho.

English Translation

Once we've rejected Ho we look to the alternative to state our conclusion

At a significance level of 0.025, there is enough evidence to support the claim that the average "heavy metal" concentrations in lake trout has decreased

Interpreting Computer Output

- A soil biologist is interested in determining the erosion rate for a 10 mile section of hillside. A significant amount of money has been spent over the past 4 years to upgrade the landscaping and she would like to determine if the erosion rate has decreased from its historical value of 5.5 tons per year. Erosion data for the last 4 years are: 2.5, 3.1, 6.2, 2.9 tons

Statistical Tests Example 2

- Step 1
  - Ho: $\mu \geq 5.5$
  - Ha: $\mu < 5.5$
- Is this a 1 tail or 2 tailed test? Why?
- Step 2: Let $\alpha = 0.05$
- Steps 3-5 at this point you can use the computer output to complete the test by hand or we can use the P-value

Interpreting Computer Output

Descriptive Statistics

Variable	N	Mean	Median	StDev	SE Mean
Erosion	4	3.675	3.000	1.702	0.851

Variable	Min	Max	Q1	Q3
Erosion	2.500	6.200	2.600	5.425

Interpreting Computer Output

Test of mu = 5.500 vs mu < 5.500

Variable	N	Mean	StDev	SE Mean	T	P
Erosion	4	3.675	1.702	0.851	-2.14	0.061

P-values and the alpha level

- In this example the P-value = 0.061 and the significance level is 0.05 so we would fail to reject Ho.
- The English interpretation is still written in terms of Ha:
  - At α = 0.05 there is not enough evidence to conclude that the erosion rates have decreased over the past 4 years.

Statistical Tests Involving Paired Data

Hypothesis test procedures can be applied to any population parameter of interest. In these next few slides we investigate the hypothesis test setting for paired data.

This is often referred to as the *"paired t-test"*

Setting

A physical fitness program is designed to increase a person's upper body strength. To determine the effectiveness of this program a SRS of 31 members of a health club was selected and asked to do as many push-ups as possible in 1 minute. After 1 month on the program the participants were once again asked to do as many push-ups as possible in 1 minute. These values were recorded and the difference (After - Before) was computed. (Use α = 0.05)

Sample data (Partial Listing)

Subject	Before	After	Difference
1	28	32	4
2	34	32	-2
3	28	42	14
4	60	64	4
5	20	41	21
6	25	33	8
7	32	49	17
8	19	32	13
9	29	50	21

- What is the interpretation for subject 2's results?

Step 1: The Hypotheses

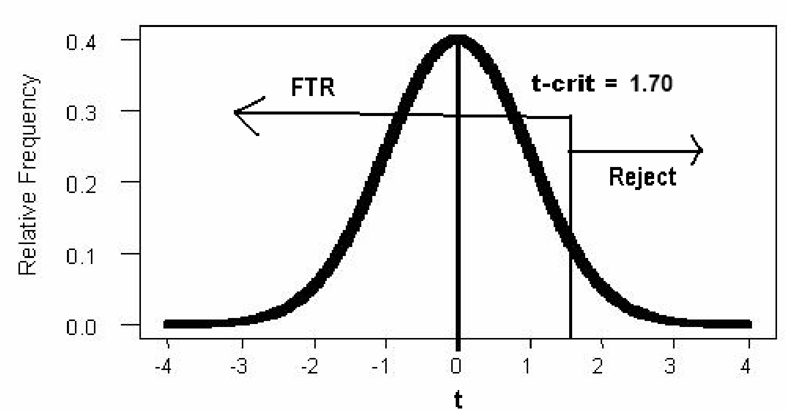
- The program administrators would like to show that this fitness program is effective. This means that a person should be able to do more push-ups after being in the program. Hence:
- Ho: $\mu_d \leq 0$ What does this mean?

  Ha: $\mu_d > 0$ What does this mean?

Step 2: Significance level

- Based on the alternative hypothesis this is a right tailed test
- $t_{crit} = 1.70$

Step 3: The decision Graphic



Descriptive Statistics

- Descriptive Statistics

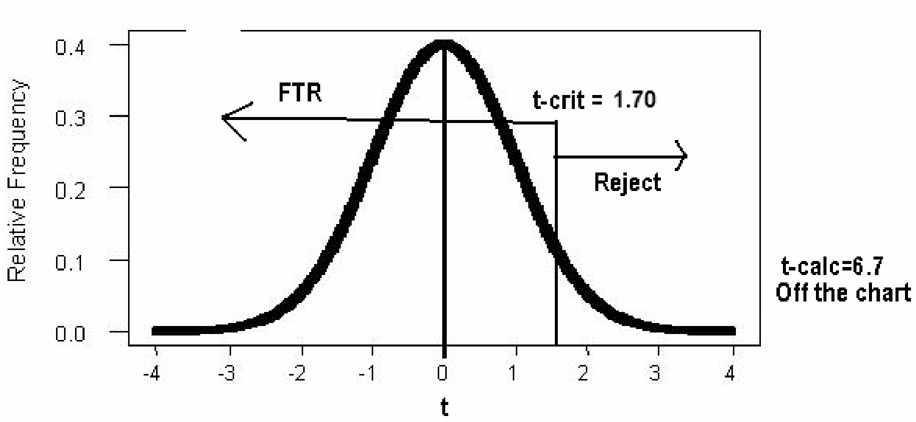
Variable	N	Mean	Median	Tr Mean	StDev	SE Mean
diff	31	9.17	8.00	8.97	8.06	1.45

Variable	Min	Max	Q1	Q3
diff	-8.00	32.00	3.00	15.00

Step 4: The Calculated Value

$$t_{calc} = \frac{\bar{d} - \mu_d}{s_d / \sqrt{n}} = \frac{9.17 - 0}{8.06 / \sqrt{31}} = 6.71$$

Step 5: The decision graphic



Steps 6 and 7

- Since t_{calc} = 6.7 lies in the rejection region we reject Ho
- We can conclude that, at a significance level of 0.05, the fitness program does, indeed, increase upper body strength.

Hypothesis tests for two means

- As you might expect a statistical test can be done to estimate whether the means from two independent populations are the same
  - We can also test to see if one population average is more (or less) than the other
- We use the same 7-step process that has been discussed extensively in previous lectures

Setting

An environmental chemist would like to estimate whether or not a certain formulation of synthetic fiber disintegrates more the longer it's buried. She randomly selects a SRS of 40 swatches of fabric and assigns them to be buried for either 4 weeks or 24 weeks. At the end of the prescribed time she digs up the fabric and measures their breaking strength

Summary Statistics

Descriptive Statistics

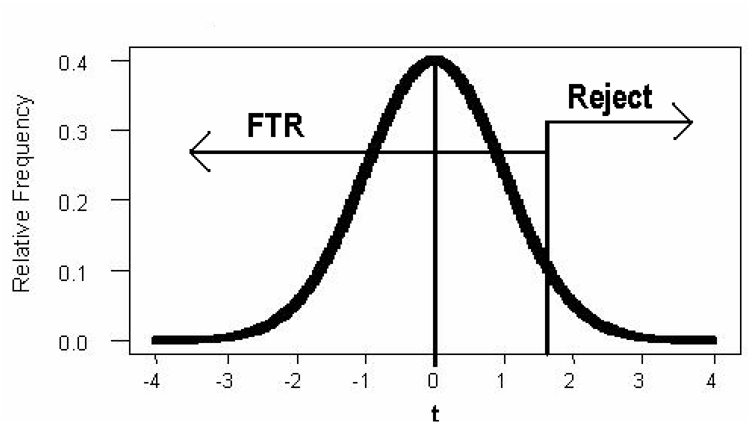
Variable	N	Mean	Median	StDev	SE Mean
4Weeks	20	116.01	116.27	4.61	1.03
24Weeks	20	91.35	88.86	9.11	2.04

Variable	Minimum	Maximum	Q1	Q3
4Weeks	106.60	126.10	113.94	119.71
24Weeks	79.13	111.02	84.72	98.34

Steps 1 and 2

- Ho: $\mu_4 - \mu_{24} \leq 0$ | Ho: $\mu_4 \leq \mu_{24}$
  Ha: $\mu_4 - \mu_{24} > 0$ | Ha: $\mu_4 > \mu_{24}$
- Let $\alpha = 0.05$; df = 19 + 19 = 38
- t = 1.68 (closest df = 40)

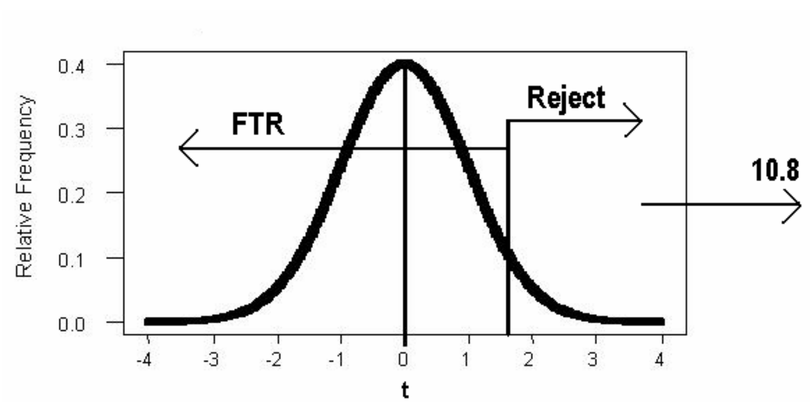
Step 3: The decision graphic



Step 4: The Test Statistic

$$
\begin{aligned}
t_{calc} &= \frac{\overline{x}_4 - \overline{x}_{24}}{\sqrt{\frac{s_4^2}{n_4} + \frac{s_{24}^2}{n_{24}}}} \\
&= \frac{116.01 - 91.35}{\sqrt{\frac{4.61^2}{20} + \frac{9.11^2}{20}}} \\
&= 10.8
\end{aligned}
$$

Step 5: The decision graphic



Steps 6 and 7

- Obviously, the statistical decision is to reject Ho
- At alpha = 0.05 we can conclude that fabric buried for 4 weeks is stronger than fabric buried for 24 weeks. This means that this fabric disintegrates more the longer it’s buried.

Computer Output

Two Sample T-Test and Confidence Interval

```
Two sample T for 4 Weeks vs 24 Weeks
          N      Mean     StDev   SE Mean
4Weeks   20    116.01      4.61       1.0
24Weeks  20     91.35      9.11       2.0

95% CI for mu 4Weeks - mu 24Weeks: ( 20.0,  29.3)
T-Test mu 4Weeks = mu 24Weeks (vs >):
 T = 10.81  P = 0.0000  DF = 28
```


Hypothesis testing: 2-proportions

A study was done that investigate the drop-out rates of college freshman that attended a 4-year private vs 4-year public colleges. In a SRS of 430 freshman in private colleges, 138 dropped out after the 1^{st} year. Of 500 public college freshman 135 dropped out. Is there evidence at $\alpha = 0.05$ to conclude that the private colleges have greater drop-out rates than the public institutions?

Step 1: Ho and Ha

Ho: $P_{pr} \leq P_{pb}$
Ha: $P_{pr} > P_{pb}$

In English this pair of hypotheses says that the proportion of first-year drop-outs in public colleges is higher than that for private colleges (Ha); Ho says that the drop-out rate for private schools is at most, as large as that in public schools.

Step 1: The Math version

Convert to the math form.

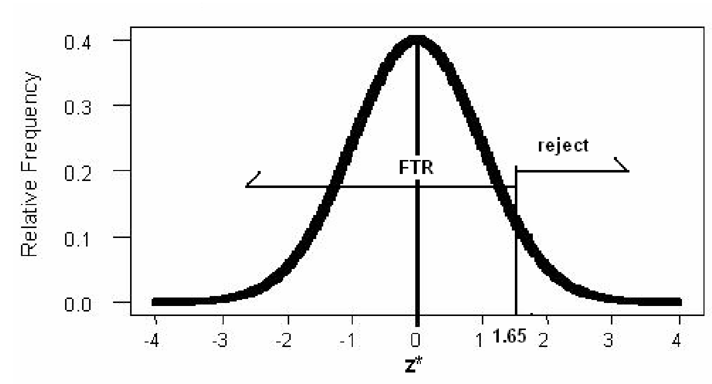
Ho: $P_{pr} - P_{pb} \leq 0$
Ha: $P_{pr} - P_{pb} > 0$

We do this so that we've got a test value on the right hand side of the equation.

Step 2: The critical Value

- The critical value is a z-value because we're using a statistical technique on proportions
- At a significance level of 0.05 z-crit = 1.65

Step 3: The decision graphic



The test statistic

- As usual the test statistic has its own formula.
- It is a z-calc

$$z_{calc} = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\overline{p}\overline{q} \bullet \left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

The test statistic

Where $\hat{p}_1$ and $\hat{p}_2$ are the sample proportions. We already know how to calculate these.

The n_1 and n_2 terms represent the sizes of each of the samples taken from each population

And $\overline{p}$ is what is called the pooled sample proportion

The Pooled Sample Proportion

- The pooled sample proportion is found by doing the following arithmetic:

$$\bar{p} = \frac{\text{Number of outcomes of interest}}{\text{Number of trials}}$$

The pooled sample proportion

- In this example the pooled sample proportion is:

$$\bar{p} = \frac{138+135}{430+500} = \frac{273}{930} = 0.294$$

The test statistic

Based on our sample results and the pooled sample proportion we get the following:

$$\hat{p}_{pb} = \frac{135}{500} = 0.270$$

$$\hat{p}_{pr} = \frac{138}{430} = 0.321$$

$$\bar{p} = 0.294$$

The test statistic

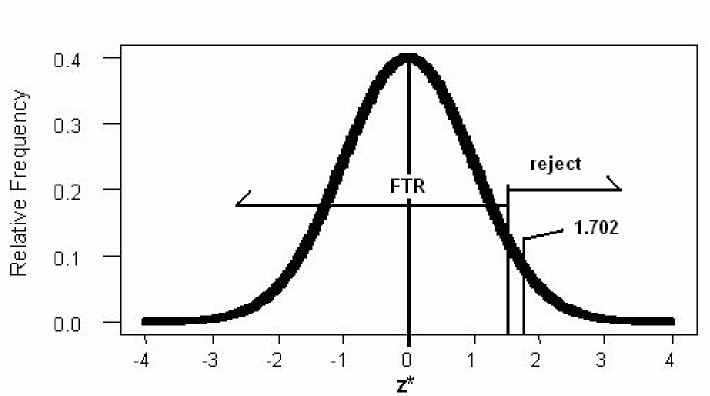
Now it's just plug-n-chug

$$z_{calc} = \frac{0.321 - 0.270}{\sqrt{(0.294)(0.706)\left(\frac{1}{430} + \frac{1}{500}\right)}}$$

$$= \frac{0.051}{\sqrt{0.00089783}}$$

$$= 1.702$$

The decision graphic



Steps 6 and 7

- It's close but z-calc is in the rejection region so we reject Ho
- We can conclude that the drop-out rate for private 4-year colleges is higher than that for public institutions at a significance level of 5%.

Interpreting computer output

Test and CI for Two Proportions

```
Sample       X      N  Sample p
1          138    430  0.320930
2          135    500  0.270000

Estimate for p(1) - p(2):  0.0509302
Test for p(1) - p(2) = 0 (vs > 0):
Z = 1.70  P-Value = 0.045
```


P-values and the alpha level

- The general rule for hypothesis testing is:

If the p-value is less than the specified alpha level then you reject Ho.

- In this case the sample data do not favor the null hypothesis.
- The p-value (0.045) is less than the specified alpha (0.05) level so Ho is rejected.

Tests of association

- When we studied association between two quantitative variables (recall extrusion temperature vs strength) we used regression techniques.
- We can't construct mathematical models, like regression lines for qualitative variables so a different technique is required.
- Statistical techniques that "measure" the relationship between categorical variables are called tests of association.

Tests of association: Philosophy

- The statistical test that is conducted is called a chi-squared test (χ^2). Each cell discrepancy contributes to the overall value for Chi-square.
- These discrepancies are based on the expected cell counts. The assumption underlying the cell counts is that the null hypothesis is **true**

Tests of association: Philosophy

- If the cell discrepancies are small then this supports the null hypothesis. That is, we're getting what we would expect to get if the categories are independent.
- If the cell discrepancies are large then this supports the alternate hypothesis which says that the categories are not independent.

Tests of association: Setting

A social scientist is interested in determining if there is a relationship between how people voted in a recent election and their relative income bracket. The two study variables are vote type (democrat/ republican) and income bracket (lower/ middle/ upper) . A contingency table that summarizes these results is presented on the next slide.

Contingency table for election results by income

	Dem.	Rep.	Total
Lower	511	202	713
Middle	387	401	788
Upper	196	303	499
Total	1094	906	2000

Table of row percents for election results

	Dem.	Rep.	Total
Lower	0.717	0.283	1.00
Middle	0.491	0.509	1.00
Upper	0.393	0.607	1.00

Does there appear to be an association between a person's economic status and the way they voted in the last election?

The Hypotheses

The null/alternate hypotheses for tests of association are written.

Ho: There is no association between the row variable and the column variable

Ha: There is an association between the row and column variables

Expected Frequencies

	Dem.	Rep.	Total (R_i)
Lower	511 (390)	202 (323)	713
Middle	387 (431)	401 (357)	788
Upper	196 (273)	303 (226)	499
Total (C_j)	1094	906	2000 (n)

To get expected frequencies use the formula:

$E_{ij} = (R_i)(C_j)/n$

Note that the sum of the E_{ij}'s = n

Expected Frequencies

The sum of all E_{ij}'s = n

The row sum for any row of E_{ij} is the corresponding marginal total

The column sum for any column of E_{ij} is the corresponding marginal total

Note that all E_{ij}'s ≥ 5

Expected Frequencies Example

	Dem.	Rep.	Total (R_i)
Lower	511	202	713
Middle	387		788
Upper		303	499
Total (C_j)	1094	906	2000 (n)

$E_{21} = (R_2)(C_1)/n$

$E_{21} = (788)(1094)/2000 = 431$

Cell Discrepancies

	Dem.	Rep.	Total (R_i)
Lower	511 (390)	202 (323)	713
Middle	387 (431)	401 (357)	788
Upper	196 (273)	303 (226)	499
Total (C_j)	1094	906	2000 (n)

The cell discrepancy is given the symbol d_{ij}

$$d_{ij} = \frac{(\text{Observed - Expected})^2}{\text{Expected}}$$

Cell Discrepancies

	Dem.	Rep.	Total (R_i)
Lower	511 (390)	202 (323)	713
Middle	387 (431)	401 (357)	788
Upper	196 (273)	303 (226)	499
Total (C_j)	1094	906	2000 (n)

The cell discrepancy is given the symbol d_{ij}

$$d_{ij} = \frac{(387 - 431)^2}{431} = 4.49$$

Cell Discrepancies

Cell	Observed	Expected	Discrepancy
C11	511	390	37.5
C21	387	431	4.5
C31	196	273	21.7
C12	202	323	45.3
C22	401	357	5.4
C32	303	226	26.2

The Chi-Square Test results

```
Chi-Square Test
Expected counts are printed below observed
   counts
          Demo     Repb    Total
     1     511      202      713
➔  expected  390.01   322.99
     2     387      401      788
➔  expected 431.04   356.96
     3     196      303      499
➔  expected 272.95   226.05
Total     1094      906     2000
```


The p-value

- Chi-Sq =

  37.533 + 45.321 + 4.499 + 5.432 + 21.695 + 26.197 = 140.678
- DF = 2,

  (DF = (N_row -1)(N_col -1)
- P-Value = 0.000

Recall P-values and Hyp. Tests

- If $p < 0.01$ reject Ho
- If $p > 0.10$ Do not reject Ho
- If p is in between these two values then look at the significance level. If p is less than alpha then reject otherwise you fail to reject the null hypothesis.
- In this instance p = 0.000. This is very small. Which means that there is a very small probability that we would get these results by chance alone. Hence, reject Ho and conclude that the variables are related in some fashion.