Hypothesis Testing

Statistical procedures for addressing research questions involves formulating a concise statement of the hypothesis to be tested.

The hypothesis to be tested is referred to as the *null hypothesis* (abbreviated H_o) because it is a statement of *no difference.*

Hypothesis testing starts with the assumption that the null hypothesis is true… that there is/are no difference(s).

Along with the null hypothesis we must also state an alternate hypothesis (abbreviated H_a).

The alternate hypothesis is a statement that a difference exists.

If a null hypothesis is rejected, then we *tentatively* accept the alternate hypothesis and conclude that there is a difference.

Why is the null hypothesis the one that is tested?

Think about it this way: we only have to find one instance in which the null hypothesis is not true (false) in order to be able to reject it.

Conversely, we would have to continue to test the alternate hypothesis in order to be able to accept it.

In other words we would have to test all possibilities since the alternate hypothesis can only be proven correct if *all* possible tests are performed.

The moral of the story:

It is easier to prove a null hypothesis incorrect than to prove an alternate hypothesis correct.

Example Hypotheses:

H_o : There is no difference in eye color between the two groups.

H_a : There is a difference in eye color between the two groups.

IMPORTANT:

In ***ALL*** cases, if your calculated probability or probability range is ***less*** than 0.05, then you ***REJECT*** the null hypothesis.

If your calculated probability or probability range is ***greater*** than 0.05, then you ***ACCEPT*** the null hypothesis.

When a statistical test is performed there are two distinct, but related results:

- Test statistic: this is the value that is calculated.

- Probability: this is the probability of having that particular test statistic given the characteristics of the data set.

These characteristics are values such as the mean, standard deviation, sample size, etc…

Which characteristics are important depends on the specific test.

The results of any statistical test (e.g. one where you are testing a null hypothesis) must be stated in a concise summary statement. This statement should include a summary of the findings, the test that was performed, *the alpha level used IF different than 0.05*, the statistical results, and the probability.

A statement of direction when a 1-tailed tested is used.

The test performed.

There is significantly less snow pack on the south side of Volcán Tancítaro ($t_{4.15}$, $p < 0.005$).

The test statistic.

Concise summary of the test results.

The probability of the test statistic.

<u>Alpha Level (α)</u>

The alpha level is the probability of committing a Type I error.

- A true null hypothesis that is incorrectly rejected.
- Also called the *significance* level.
- It is essentially a ‘false positive’.

By convention we typically use 0.05 or 0.01 (5% or 1%) as our alpha level.

<u>Beta Level (β)</u>

The beta level is the probability of committing a Type II error.

- A false null hypothesis we fail to reject.
- It is essentially a ‘false negative’.

By convention this value is not specified.

Type 1 and 2 Errors

	If H_o is true	***If H_o is false***
And H_o is rejected:	Type I Error	No Error
And H_o is not rejected:	No Error	Type II Error

H_o: There is no significant difference in dissolved oxygen between pond 1 and pond 2.

Ho: There is no difference in survival rates between *Test Drug A* and a placebo.

In reality this is true... There really is no difference in survival.

A Type I error would occur if we incorrectly reject this true null hypothesis.

- The drug would go to market.
- People would take the drug expecting to survive.
- They would die since the drug has no effect.

This is bad...

Ho: There is no difference in survival rates between *Test Drug A* and a placebo.

In reality this is false... There really is difference in survival.

A Type II error would occur if we incorrectly accept this false null hypothesis.

- The drug would not go to market since our statistics showed it did not increase survival.
- People would never be given this drug.
- They would die since the drug has a positive effect but we said it did not.

This is also bad...

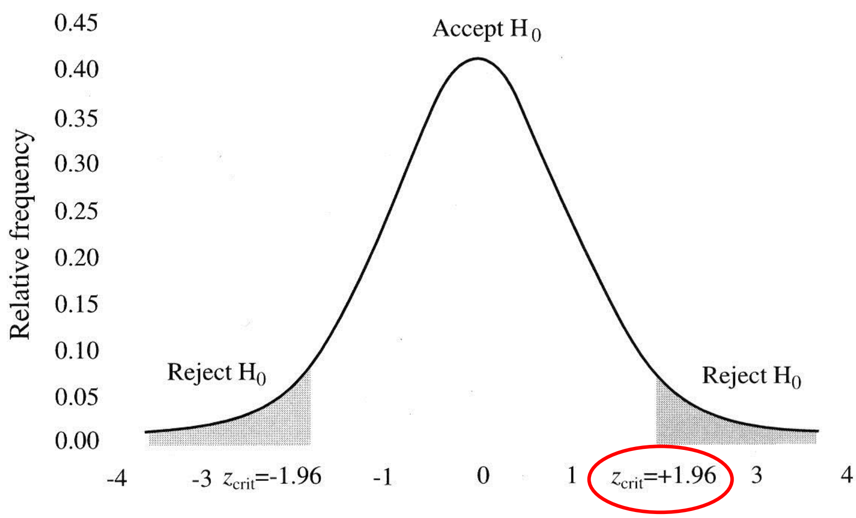
A Type I error accepts an alternate hypothesis when the results can be *attributed to chance*.

- So in effect we are stating that there is a difference when none actually exists.

A Type II error is only an error in the sense that we fail to correctly reject a false null hypothesis.

- It is not an error in the sense that an incorrect conclusion was drawn since no conclusion is drawn when the null hypothesis is accepted. It is considered the lesser of two evils.

Alpha level of 0.05



An α of 0.05 is equal to ± 1.96 sd from the mean.

One and Two-tailed Tests

- Two-tailed statistics test for difference *only*.

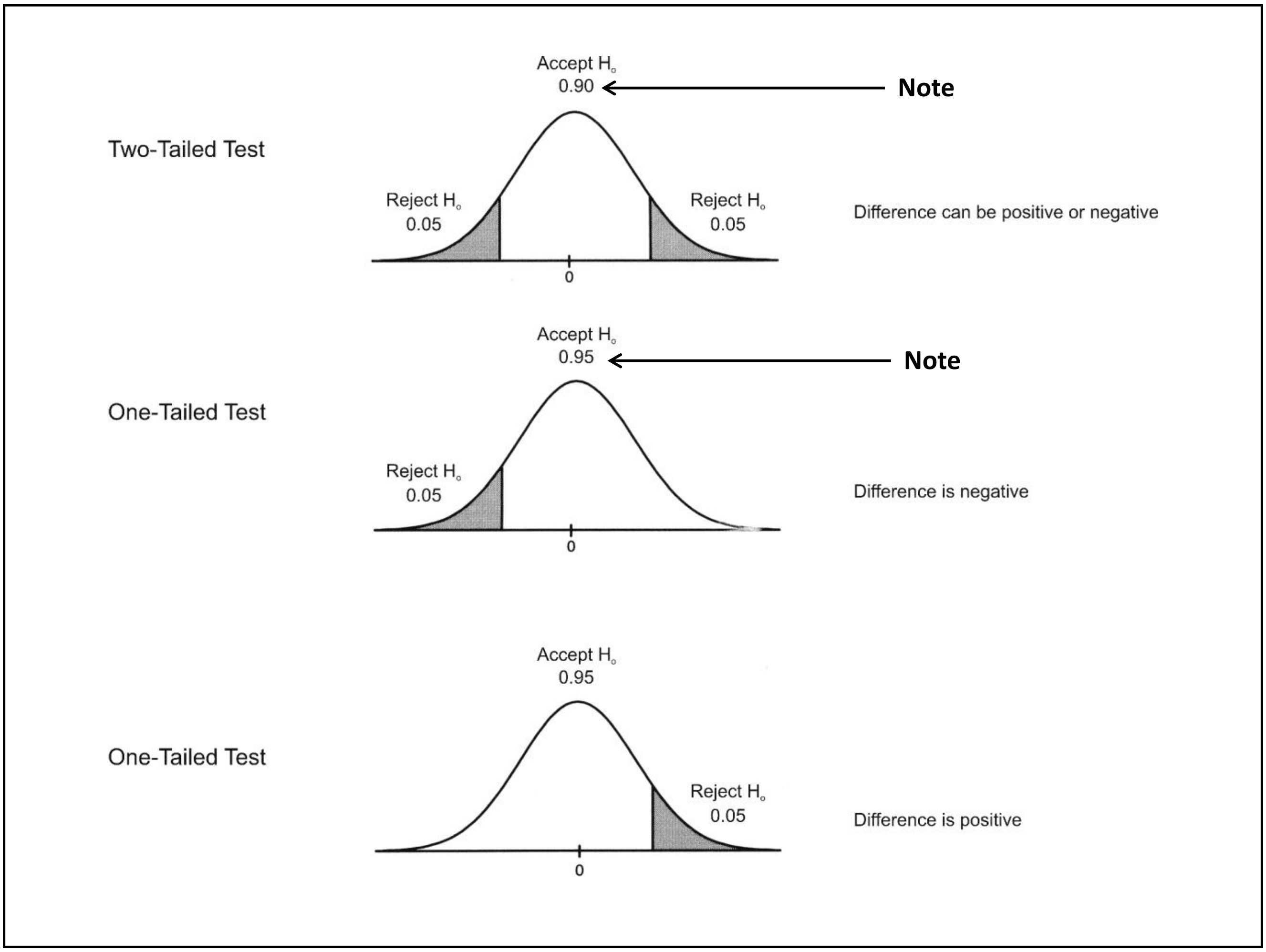
Ho: The rate of erosion at location A ***is not*** significantly different than the rate of erosion at location B.

Ha: The rate of erosion at location A ***is*** significantly different than the rate of erosion at location B.

- One-tailed statistics test for difference *and* direction.

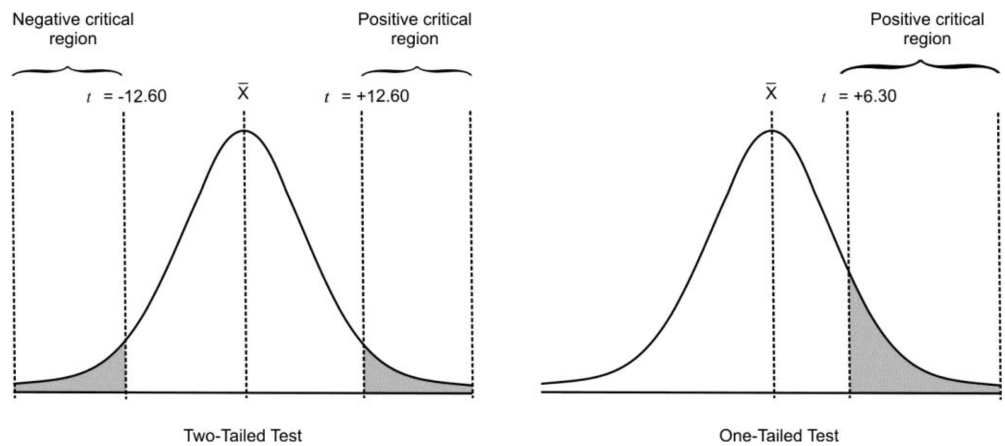
Ho: The rate of erosion at location A ***is not*** significantly ***greater*** than the rate of erosion at location B.

Ha: The rate of erosion at location A ***is*** significantly ***greater*** than the rate of erosion at location B.



Accept H$_o$
0.90
Note
Two-Tailed Test
Reject H$_o$
0.05
Reject H$_o$
0.05
Difference can be positive or negative
0
Accept H$_o$
0.95
Note
One-Tailed Test
Reject H$_o$
0.05
Difference is negative
0
Accept H$_o$
0.95
One-Tailed Test
Reject H$_o$
0.05
Difference is positive
0

Two-tailed test have the critical region split between positive and negative sides of the distribution. One-tailed tests do not. This is reflected in the table values (larger for two-tailed, smaller for one-tailed).



Earlier research suggested that the average house value in York was lower than that in Lancaster.

Data for the average house value for each block group in downtown York and Lancaster were gathered from the census.

A two-sample t test was performed to determine whether housing values were lower in York than in Lancaster.

Since we have *a priori* (prior) knowledge of the direction of the difference (e.g. housing values are *less* in York) we would use a *one-tailed test*.

H_o : Housing values in York are not significantly less than those in Lancaster.

H_a : Housing values in York are significantly less than those in Lancaster.

Note how the direction of the difference is stated in the hypotheses.

The test we will use is called a T-Test:

$$t = \frac{\overline{X}_1 - \overline{X}_2}{s_{\overline{X}_1 - \overline{X}_2}} \quad where \quad s_{\overline{X}_1 - \overline{X}_2} = \sqrt{\frac{s_p^2}{n_1} + \frac{s_p^2}{n_2}} \qquad s_p^2 = \frac{SS_1 + SS_2}{v_1 + v_2}$$

Housing Value ($)	
York	Lancaster
25368	49465
37045	37500
47500	53055
26785	48125
41493	45000
32864	52946
26140	

$\alpha = 0.05$

$n_1 = 7 \quad n_2 = 6$

$df = (n_1 + n_2 - 2) = (7 + 6 - 2) = 11$

$v_1 = 7 - 1 = 6$

$v_2 = 6 - 1 = 5$

$$\bar{X}_{York} = \frac{25368 + 37045 + 47500 + 26785 + 41493 + 32864 + 26140}{7} = 33885$$

$$\bar{X}_{Lancaster} = \frac{49465 + 37500 + 53055 + 48125 + 45000 + 52946}{6} = 47682$$

$$SS_{York} = (25368 - 33885)^2 + (37045 - 33885)^2 + (47500 - 33885)^2 + (26785 - 33885)^2 + (41493 - 33885)^2 + (32846 - 33885)^2 + (26140 - 33885)^2 = 437212244$$

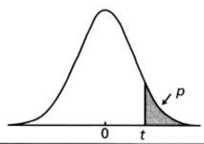
$$SS_{Lancaster} = (49465 - 47682)^2 + (37500 - 47682)^2 + (53055 - 47682)^2 + (48125 - 47682)^2 + (45000 - 47682)^2 + (52946 - 47682)^2 = 2444393535$$

$$s_p^2 = \frac{437212244 + 2444393535}{6 + 5} = \frac{2881605779}{11} = 261964161.7$$

$$s_{X_1 - X_2} = \sqrt{\frac{261964161.7}{7} + \frac{261964161.7}{6}} = \sqrt{37423451.7 + 43660693.6} = \sqrt{81084145.3} = 9004.7$$

$$t = \frac{33885 - 47682}{9004.7} = -1.532 \qquad \text{(ignore the sign when using the table)}$$

$$t_{Critical} = 1.796 \qquad \textit{Since } 1.532 < 1.796, \text{ accept } \mathrm{H}_0$$



df	p (one-tailed probabilities)				
	.10	.05	.025	.01	.005
1	3.078	6.314	12.706	31.821	63.657
2	1.886	2.920	4.303	6.965	9.925
3	1.638	2.353	3.182	4.541	5.841
4	1.533	2.132	2.776	3.747	4.604
5	1.476	2.015	2.571	3.365	4.032
6	1.440	1.943	2.447	3.143	3.707
7	1.415	1.895	2.365	2.998	3.499
8	1.397	1.860	2.306	2.896	3.355
9	1.383	1.833	2.262	2.821	3.250
10	1.372	1.812	2.228	2.764	3.169
11	1.363	$\boxed{1.796}$	2.201	2.718	3.106
12	1.356	1.782	2.179	2.681	3.055
13	1.350	1.771	2.160	2.650	3.012
14	1.345	1.761	2.145	2.624	2.977
15	1.341	1.753	2.131	2.602	2.947
16	1.337	1.746	2.120	2.583	2.921
17	1.333	1.740	2.110	2.567	2.898
18	1.330	1.734	2.101	2.552	2.878
19	1.328	1.729	2.093	2.539	2.861
20	1.325	1.725	2.086	2.528	2.845
21	1.323	1.721	2.080	2.518	2.831
22	1.321	1.717	2.074	2.508	2.819
23	1.319	1.714	2.069	2.500	2.807
24	1.318	1.711	2.064	2.492	2.797
25	1.316	1.708	2.060	2.485	2.787
26	1.315	1.706	2.056	2.479	2.779
27	1.314	1.703	2.052	2.473	2.771
28	1.313	1.701	2.048	2.467	2.763
29	1.311	1.699	2.045	2.462	2.756
30	1.310	1.697	2.042	2.457	2.750
40	1.303	1.684	2.021	2.423	2.704
60	1.296	1.671	2.000	2.390	2.660
120	1.289	1.658	1.980	2.358	2.617
∞	1.282	1.645	1.960	2.326	2.576

Adapted from Table III of Fisher and Yates (1974).

$t_{Critical} = 1.796$

Since $1.532 < 1.796$ we accept H_o.

Calculated t value. **Critical value from the table.**

The calculated t value (1.532) is then compared to the critical t value (1.796).

- IF it is higher then we REJECT H_o.
- IF it is lower then we ACCEPT H_o.

The critical values are either taken from a table or calculated in SPSS.

Our HIGH POWER summary statement would then be:

Since in this class we will, by convention, use an alpha level of 0.05, our high power summary statement would read:

The housing values in York were not significantly less than the housing values in Lancaster ($t_{1.532}$, $0.10 > p > 0.05$).

Include the alpha level in the summary statement ONLY if it is different than 0.05. However, it must be stated somewhere.

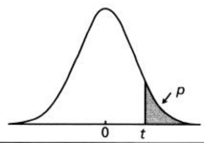
The housing values in York were not significantly less than the housing values in Lancaster ($t_{1.532}$, $0.10 > p > 0.05$).

How do we determine this?

This is called a *probability range*. If we are using a table it is rare that our calculated value will match the table values exactly.

The best we can do is state that our calculated value fell between two probabilities from table.

1.532 would fall about here on the table for a df of 11 at an alpha level of 0.05. So our probability range is between 0.10 and 0.05.



df	p (one-tailed probabilities) .10	.05	.025	.01	.005
1	3.078	6.314	12.706	31.821	63.657
2	1.886	2.920	4.303	6.965	9.925
3	1.638	2.353	3.182	4.541	5.841
4	1.533	2.132	2.776	3.747	4.604
5	1.476	2.015	2.571	3.365	4.032
6	1.440	1.943	2.447	3.143	3.707
7	1.415	1.895	2.365	2.998	3.499
8	1.397	1.860	2.306	2.896	3.355
9	1.383	1.833	2.262	2.821	3.250
10	1.372	1.812	2.228	2.764	3.169
11	1.363	1.796	2.201	2.718	3.106
12	1.356	1.782	2.179	2.681	3.055
13	1.350	1.771	2.160	2.650	3.012
14	1.345	1.761	2.145	2.624	2.977
15	1.341	1.753	2.131	2.602	2.947
16	1.337	1.746	2.120	2.583	2.921
17	1.333	1.740	2.110	2.567	2.898
18	1.330	1.734	2.101	2.552	2.878
19	1.328	1.729	2.093	2.539	2.861
20	1.325	1.725	2.086	2.528	2.845
21	1.323	1.721	2.080	2.518	2.831
22	1.321	1.717	2.074	2.508	2.819
23	1.319	1.714	2.069	2.500	2.807
24	1.318	1.711	2.064	2.492	2.797
25	1.316	1.708	2.060	2.485	2.787
26	1.315	1.706	2.056	2.479	2.779
27	1.314	1.703	2.052	2.473	2.771
28	1.313	1.701	2.048	2.467	2.763
29	1.311	1.699	2.045	2.462	2.756
30	1.310	1.697	2.042	2.457	2.750
40	1.303	1.684	2.021	2.423	2.704
60	1.296	1.671	2.000	2.390	2.660
120	1.289	1.658	1.980	2.358	2.617
∞	1.282	1.645	1.960	2.326	2.576

Adapted from Table III of Fisher and Yates (1974).

A few words on *degrees of freedom* (called *df* or *v*) …

- Think of degrees of freedom as the minimum amount of information needed to be able to determine the value of ALL of the observations.

- For example, if we know the value of n-1 observations AND the mean, we can calculate the last observation value .

Data			
5	$n = 5$	$\dfrac{\sum x_i}{5} = 5$	$\sum x_i = 5 \times 5 = 25$
7	$\bar{x} = 5$		
3		$25 - (5 + 7 + 3 + 6) = 4$	
6			
?			

Every time a statistical test is performed, you need to include the following:

1. A statement of the alpha level.
2. Null and alternate hypotheses.
3. A high power summary statement that contains:
   - The test performed.
   - The calculated test statistic.
   - The exact probability (SPSS) or probability range (Table).

NOTE: Probabilities are *ALWAYS* written in descending order:

RIGHT: $0.10 > p > 0.05$ WRONG: ~~$0.05 < p < 0.10$~~

Developing Hypotheses Based on Maps

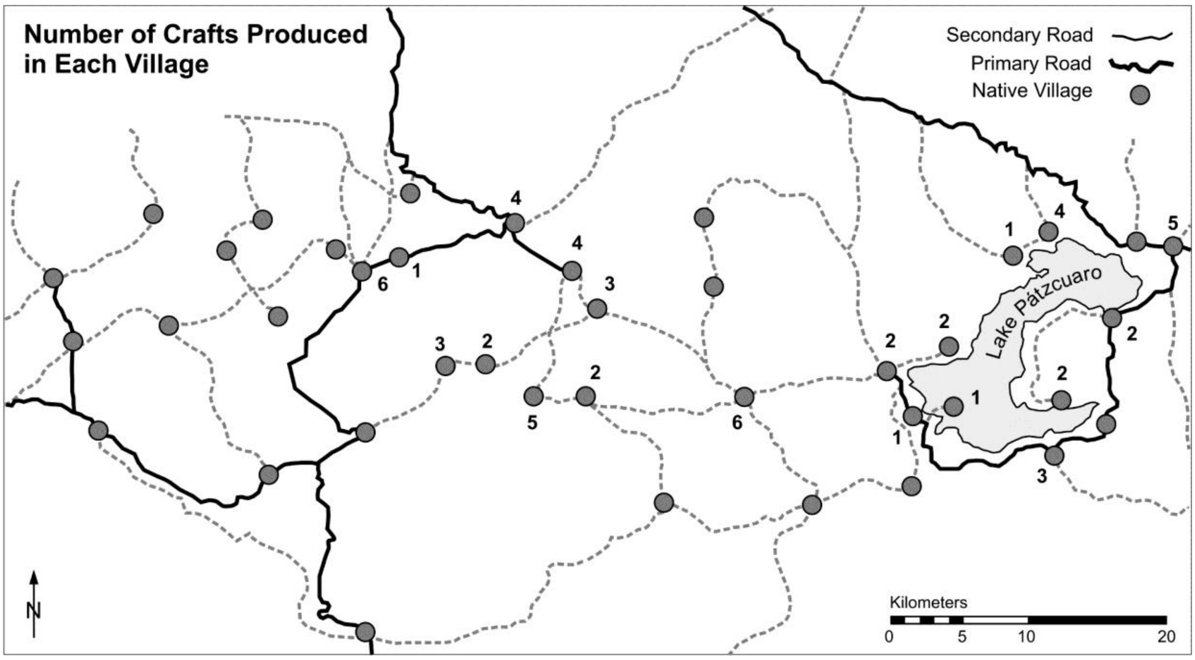
Maps can be used for:

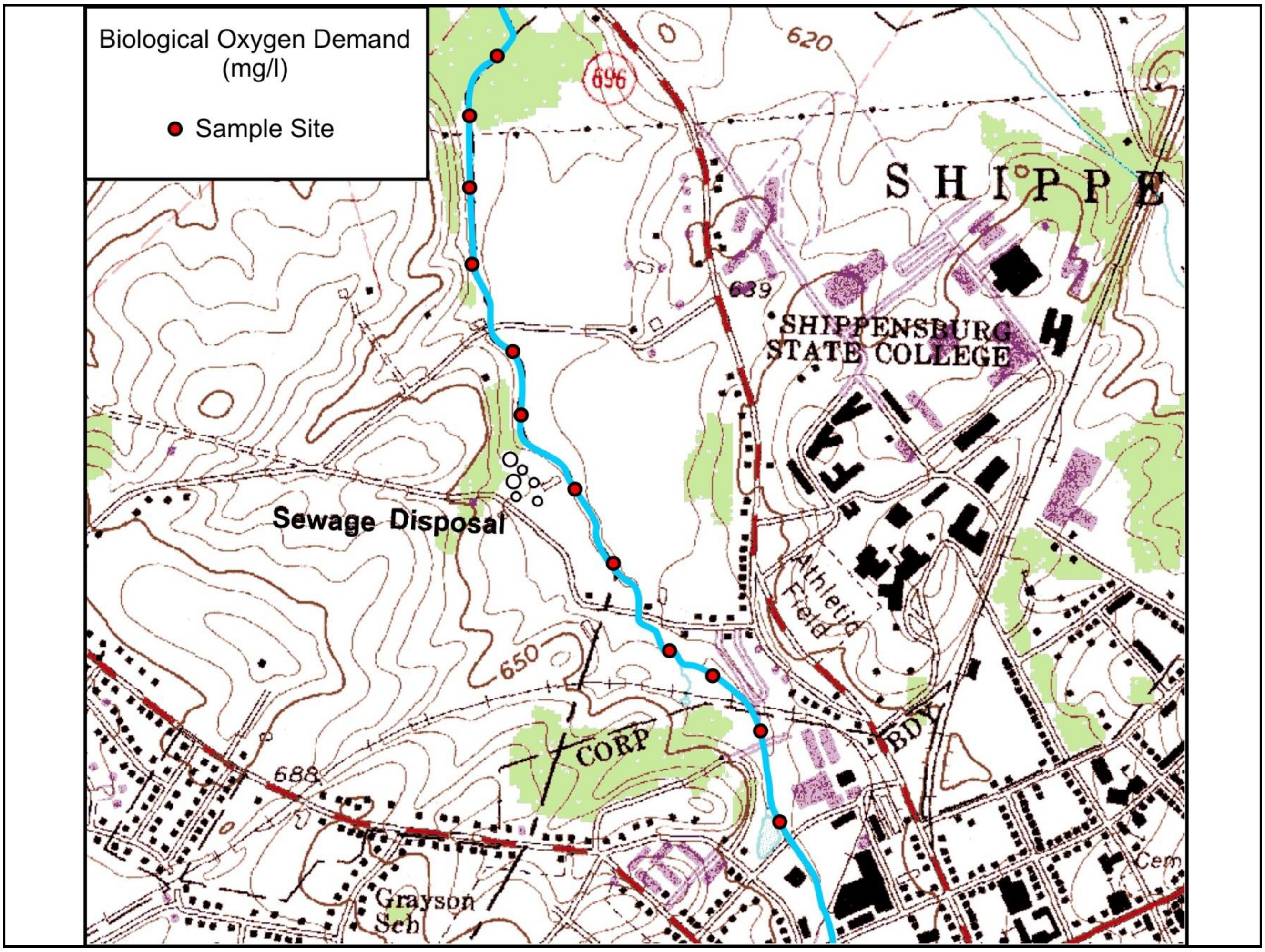
- Delineating groups
- Locating observations
- Determining measurements

Measurements can be (but are not limited to):

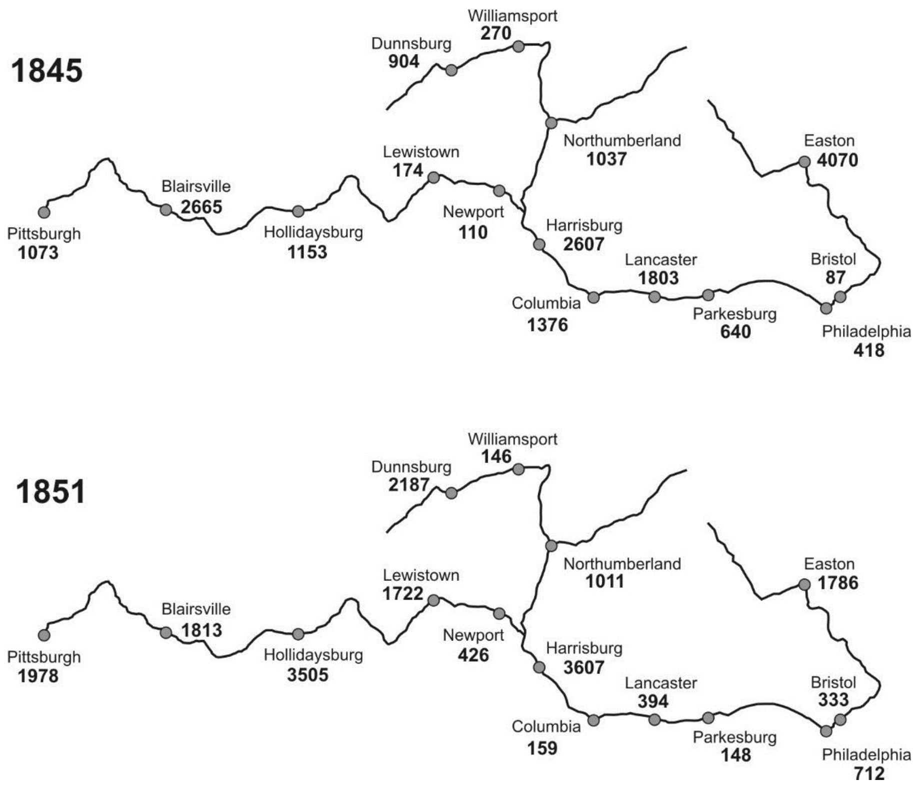
- Elevation
- Azimuth
- Aspect
- Distance
- Proximity

Develop null and alternate hypotheses based the following maps.

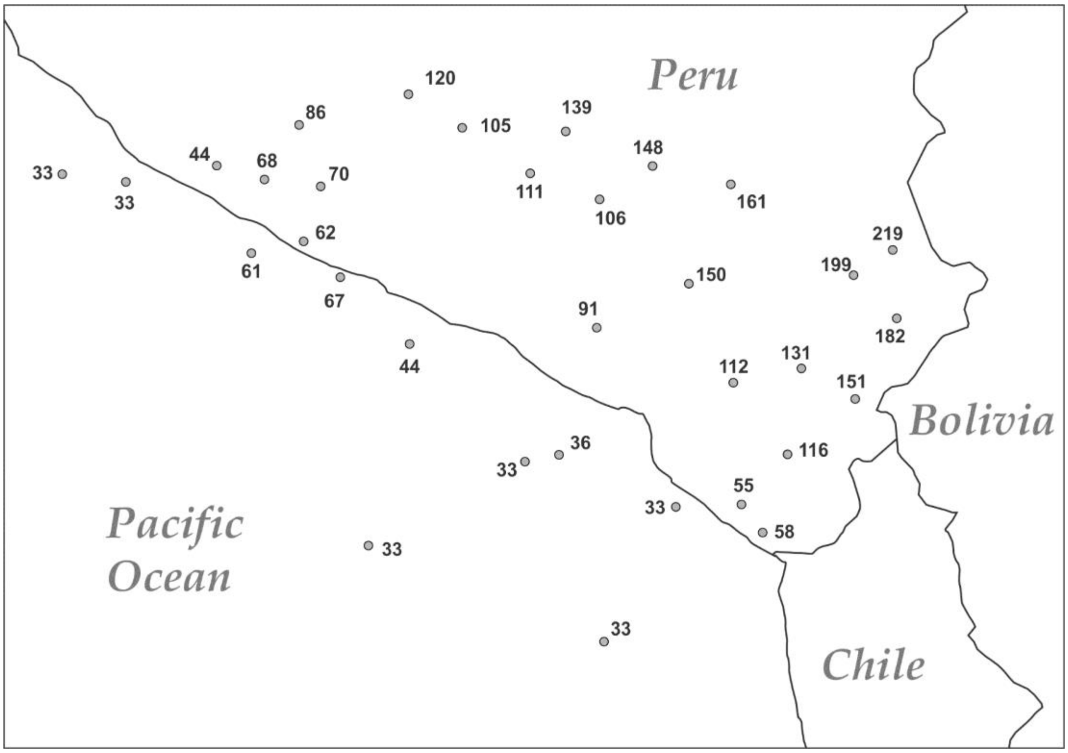


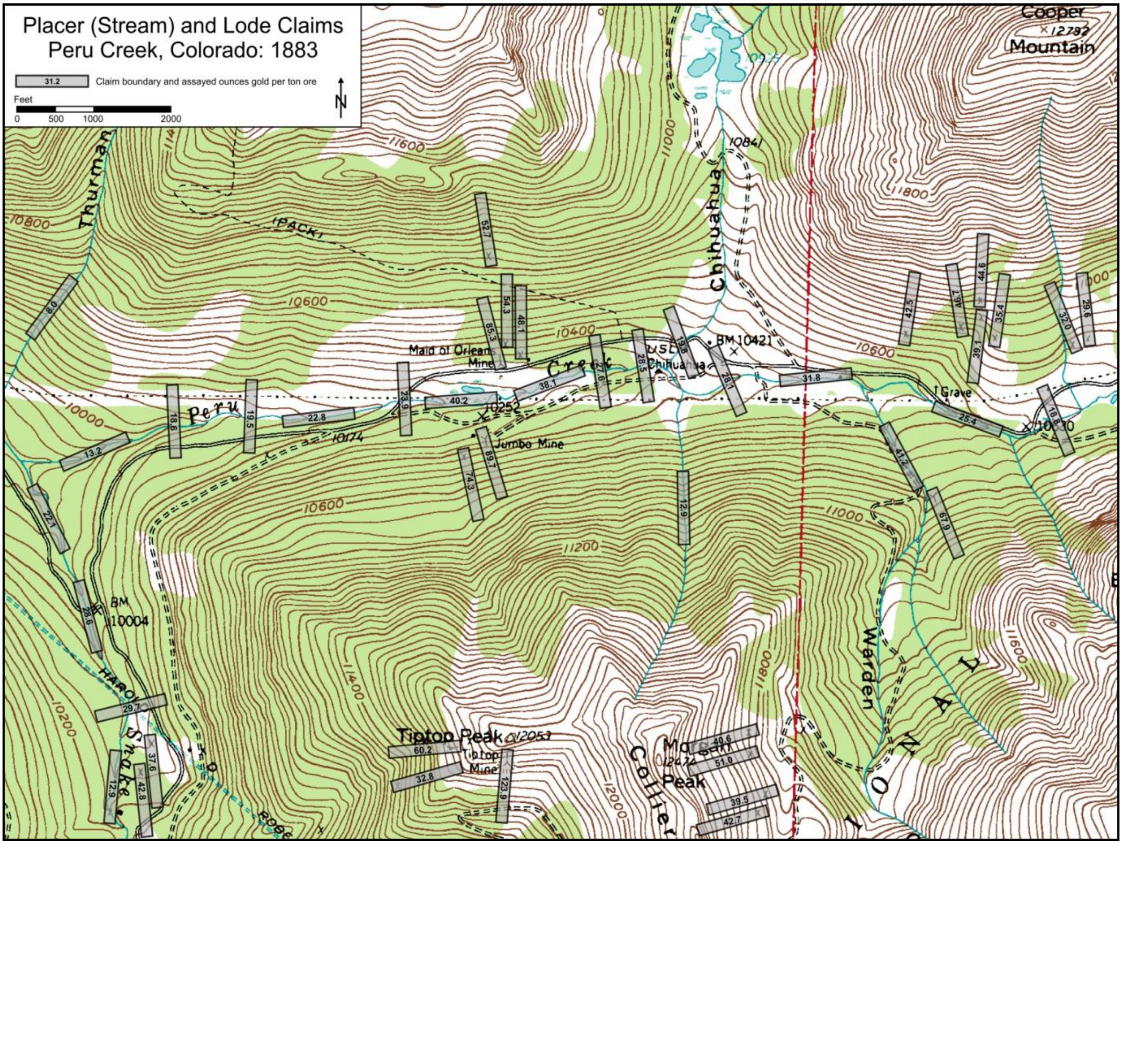

Biological Oxygen Demand (mg/l)
Sample Site
SHIPPE
SHIPPENSBURG STATE COLLEGE
Sewage Disposal
Athletic Field
CORP
BDY
Grayson Sch
Cem
620
696
639
650
688

Ton of Pig Iron Shipped via Canal



Earthquake Depths (km)




Placer (Stream) and Lode Claims
Peru Creek, Colorado: 1883
31.2
Claim boundary and assayed ounces gold per ton ore
Feet
0 500 1000 2000
N
Cooper Mountain
12792
Thurman
Chihuahua
Peru Creek
Maid of Orleans Mine
Jumbo Mine
BM 10421
BM 10004
10252
10174
10841
Grave
Tiptop Peak
12053
Tiptop Mine
Morgan Peak
12474
Warden
Snake
11600
10800
10600
10400
10000
11000
11800
11200
11400
10200
12000
52.7
54.3
48.1
85.3
8.0
18.6
19.5
13.2
22.8
23.9
40.2
38.1
27.6
28.5
19.8
28.1
31.8
42.5
46.7
44.6
35.4
39.1
32.0
29.6
18.8
25.4
41.2
67.9
12.9
89.7
74.3
22.1
28.6
29.7
37.6
42.8
60.2
32.8
123.9
40.6
51.0
39.5
42.7

Often our hypotheses are not concerned with differences between or among groups.

- For example, an association between a measured variable and a measured landscape or natural characteristic.

- e.g. Decreasing temperature with increasing altitude.

In such cases, our null and alternate hypotheses might be:

H_o: *There no association between temperature and altitude.*

H_a: *There is an association between temperature and altitude.*

Data displayed as:
Housing prices per ft^2, distance in meters to I-70