Tests of Significance: Examples

Gene Quinn

Tests of Significance

Tests of Significance, also called *inference* or *hypothesis testing* refers to the application of statistical techniques to decision making.

Tests of Significance

Tests of Significance, also called *inference* or *hypothesis testing* refers to the application of statistical techniques to decision making.

In general, Tests of Significance are used to determine whether some claim about a population is true or not.

Tests of Significance

Tests of Significance, also called *inference* or *hypothesis testing* refers to the application of statistical techniques to decision making.

In general, Tests of Significance are used to determine whether some claim about a population is true or not.

We looked at three special cases:

- The claim involves a proportion
- The claim involves a mean, and we know the population standard deviation
- The claim involves a mean, and we do not know the population standard deviation

Tests of Significance

These test involve using a *single* sample to test come claim about the population mean or proportion.

Tests of Significance

These test involve using a *single* sample to test come claim about the population mean or proportion.

A variation is the case where we have *two* samples.

Tests of Significance

These test involve using a *single* sample to test come claim about the population mean or proportion.

A variation is the case where we have *two* samples.

The claim in this case is usually one of the following:

- The two population proportions are the same
- The two population means are the same, and we know the population standard deviations
- The two population means are the same, and we do not know the population standard deviations

Tests of Significance

When preseneted a significance testing problem, we can proceed in the following manner.

Tests of Significance

When preseneted a significance testing problem, we can proceed in the following manner.

First ask whether this is a "two sample" problem, or a "one sample" problem, with a separate claim.

Tests of Significance

When preseneted a significance testing problem, we can proceed in the following manner.

First ask whether this is a "two sample" problem, or a "one sample" problem, with a separate claim.

The next steps will depend on the answer to this question.

Two Samples: Dependent

If you have two samples, you then have to ask whether the samples are paired (dependent) or independent.

Two Samples: Dependent

If you have two samples, you then have to ask whether the samples are paired (dependent) or independent.

Typical paired samples:

- Before and after measures on the same subject
- Same store sales
- Twin studies

Two Samples: Dependent

If you have two samples, you then have to ask whether the samples are paired (dependent) or independent.

Typical paired samples:

- Before and after measures on the same subject
- Same store sales
- Twin studies

In the case of paired samples, subtract the before and after measures for each subject and use that number in the analysis.

Two Samples: Dependent

Once you subtract the before and after measures for each pair, the problem is exactly like a single-sample problem.

Two Samples: Dependent

Once you subtract the before and after measures for each pair, the problem is exactly like a single-sample problem.

The claim in this case is that the population mean of the difference is zero.

Two Samples: Dependent

Once you subtract the before and after measures for each pair, the problem is exactly like a single-sample problem.

The claim in this case is that the population mean of the difference is zero.

So we can think of the paired sample situation as a single sample of differences, with the claim that the population mean of the differences is zero.

Two Samples: Independent

If you have independent (non-paired) samples, the next question is which of the following situations we have:

- Two sample proportions
- Two sample means, with the population standard deviations known
- Two sample means, with the population standard deviations unknown

Two Samples: Independent

If you have independent (non-paired) samples, the next question is which of the following situations we have:

- Two sample proportions
- Two sample means, with the population standard deviations known
- Two sample means, with the population standard deviations unknown

Careful reading of the problem should reveal which of these cases best fits the situation.

A Single Samples

If you have a single sample, the only question is which of the following three situations best fits.

- A sample proportion and a claim
- A sample mean and a claim, with the population standard deviation known
- A sample mean and a claim, with the population standard deviation unknown

A Single Samples

If you have a single sample, the only question is which of the following three situations best fits.

- A sample proportion and a claim
- A sample mean and a claim, with the population standard deviation known
- A sample mean and a claim, with the population standard deviation unknown

As before, careful reading of the problem should reveal which of these cases best fits the situation.

Example 1

A school administration maintains that SAT math scores for their district are siginficantly higher than the national average. A sample of 120 students from the graduating class had an average SAT math score of 515. SAT scores are standardized to have a mean of 500 and a standard deviation of 100. At the usual alpha level of 0.5, do the test scores for the sample support the claim of significantly higher scores for the district?

Example 1

A school administration maintains that SAT math scores for their district are siginficantly higher than the national average. A sample of 120 students from the graduating class had an average SAT math score of 515. SAT scores are standardized to have a mean of 500 and a standard deviation of 100. At the usual alpha level of 0.5, do the test scores for the sample support the claim of significantly higher scores for the district?

Category: Hypothesis testing on a mean with known standard deviation (σ known)

Most Likely Null Hypothesis: SAT math scores are *not* higher than the norm of 500.

Example 1

Conclusion: Accept the null hypothesis. The difference is not significant at the 5% level.

Example 1

Conclusion: Accept the null hypothesis. The difference is not significant at the 5% level.

The interpretation would be that the mean test scores are higher, but not so much higher that you could say that there is a less than one chance in twenty that the difference is purely due to sampling error.

You would need either a bigger difference, or a larger sample size for this difference in order to say that the average is significantly higher than the norm for the test.

Example 2

Analysis of a sample of 45 deer ticks (Ixodes scapularis) from wooded area reveals the presence of the Ehrlichiosis pathogen (the Ehrlichia chaffeensis bacteria) in 17 ticks. The CDC estimates that 18% of ticks carry the pathogen. Is the data consistent with this claim?

Example 2

Analysis of a sample of 45 deer ticks (Ixodes scapularis) from wooded area reveals the presence of the Ehrlichiosis pathogen (the Ehrlichia chaffeensis bacteria) in 17 ticks. The CDC estimates that 18% of ticks carry the pathogen. Is the data consistent with this claim?

Category: Hypothesis testing with a proportion

Most Likely Null Hypothesis: The population proportion matches the CDC estimate of 18%

Example 2

Conclusion: Reject the null hypothesis. The difference is significant at the 5% level.

Example 2

Conclusion: Reject the null hypothesis. The difference is significant at the 5% level.

Interpretation: With this sample size, if the proportion of ticks carrying the pathogen is actually 18%, there is a less than 5% chance of getting 17 infected ticks in a sample of 45.

Example 3

In a double-blind study of Major Depressive Disorder (MDD), 43 subjects are evaluated using the Beck Depression Inventory (BDI). Subjects then receive drug therapy with seratonin-reuptake inhibitors (SRIs). After 8 weeks of treatment, subjects are reevaluated with the BDI. The average difference in the before and after scaled BDI scores is -1.7 with a standard deviation of 0.8. Can we conclude that SRIs are effective in mitigating the MDD severity as measured by the BDI?

Example 3

In a double-blind study of Major Depressive Disorder (MDD), 43 subjects are evaluated using the Beck Depression Inventory (BDI). Subjects then receive drug therapy with seratonin-reuptake inhibitors (SRIs). After 8 weeks of treatment, subjects are reevaluated with the BDI. The average difference in the before and after scaled BDI scores is -1.7 with a standard deviation of 0.8. Can we conclude that SRIs are effective in mitigating the MDD severity as measured by the BDI?

Category: Hypothesis testing with paired samples

Most Likely Null Hypothesis: The drug therapy is not effective; the mean difference is zero.

Example 3

Conclusion: Accept the null hypothesis. The difference is not significant at the 5% level.

Example 3

Conclusion: Accept the null hypothesis. The difference is not significant at the 5% level.

Interpretation: With this sample size, and this standard deviation for the differences, we cannot say that the reduction in the BDI scores is significantly different from zero.

Example 4

A government agency is trying to build a case to support the assertion that ocean dumping of titanium dioxide waste in the Baltimore Canyon is resulting in detectable increases in titanium in the deep sea scallop (Placopecten magellanicus) population in the dumping area. A sample of 175 specimens from the dumping area showed titanimum levels of 13.2 parts per million with a standard deviation of 8.1. Published data indicates that the body burden of titanium in specimens from unpolluted areas averages 6.8 parts per million. Does the data support the agency's claim?

Example 4

A government agency is trying to build a case to support the assertion that ocean dumping of titanium dioxide waste in the Baltimore Canyon is resulting in detectable increases in titanium in the deep sea scallop (Placopecten magellanicus) population in the dumping area. A sample of 175 specimens from the dumping area showed titanimum levels of 13.2 parts per million with a standard deviation of 8.1. Published data indicates that the body burden of titanium in specimens from unpolluted areas averages 6.8 parts per million. Does the data support the agency's claim?

Category: Hypothesis testing on a mean with unknown standard deviation.

Most Likely Null Hypothesis: The average body burden or titanium is 6.8 parts per million in the dumping area.

Example 4

Conclusion: Reject the null hypothesis. The chances of drawing a sample of 175 with a mean of 13.2 from a population with a mean of 8.1 is less than 5%.

Example 4

Conclusion: Reject the null hypothesis. The chances of drawing a sample of 175 with a mean of 13.2 from a population with a mean of 8.1 is less than 5%.

Interpretation: Based on the sample, we would conclude that the body burden of titanium is significantly higher than 6.8 ppm in the dumping area.

Example 5

In a double-blind experiment 150 patients with falciparum malaria are divided into two groups of 75. The first group is treated with chloroquine, and the second receives ACT. After 6 weeks, there are 17 remissions in the ACT group and 8 in the chloroquine group. Does this data support the conclusion that ACT is a more effective treatment?

Example 5

In a double-blind experiment 150 patients with falciparum malaria are divided into two groups of 75. The first group is treated with chloroquine, and the second receives ACT. After 6 weeks, there are 17 remissions in the ACT group and 8 in the chloroquine group. Does this data support the conclusion that ACT is a more effective treatment?

Category: Hypothesis testing: Inference about two proportions

Most Likely Null Hypothesis: The proportion of patients that remit with ACT is the same as the proportion that remit with chloroquinine.

Example 5

Conclusion: Reject the null hypothesis. With this sample size, there is a less than 5% chance of getting a difference in the proportion of remissions as large as the one we observed.

Example 5

Conclusion: Reject the null hypothesis. With this sample size, there is a less than 5% chance of getting a difference in the proportion of remissions as large as the one we observed.

Interpretation: ACT produces significantly more remissions that chloroquinine.

Example 6

Records from a banding station indicate that in 2004, the average weight of a certain species of bird is 17.3 grams with a standard deviation of 2.0, based on 58 individual birds. In 2005, 73 individuals from this species are captured and their average weight is 19.1 grams with a standard deviation of 2.3. Is this data consistent with the claim that the average weight in the population has not changed?

Example 6

Records from a banding station indicate that in 2004, the average weight of a certain species of bird is 17.3 grams with a standard deviation of 2.0, based on 58 individual birds. In 2005, 73 individuals from this species are captured and their average weight is 19.1 grams with a standard deviation of 2.3. Is this data consistent with the claim that the average weight in the population has not changed?

Category: Hypothesis testing: Inference about two means with population standard deviation (σ) unknown

Most Likely Null Hypothesis: There is no difference in the average weights of individual birds of this species in 2004 and 2005.

Example 6

Conclusion: Reject the null hypothesis. The chances of getting two samples with means this different are less than 5% with the given sample sizes, sample means, and sample standard deviations.

Example 6

Conclusion: Reject the null hypothesis. The chances of getting two samples with means this different are less than 5% with the given sample sizes, sample means, and sample standard deviations.

Interpretation: The average weight is significantly higher in 2005 than it was in 2004.

Example 7

Mosquito traps are placed near 43 small ponds and a count of Culex species in the traps is obtained during a baseline period. At the end of the baseline period a spraying program is conducted. One week after the spraying, the traps are cleaned and a second collection period is initiated. Based on an estimate of the size of each pond, the raw counts are converted to a density of Culex mosquitoes per square foot of pond. The difference between the before and after densities is found to have a sample mean of 9.1 and a sample standard deviation of 12.0. Test whether or not the data indicates that the spraying was effective in reducing the density of Culex species mosquitoes.

Example 7

Mosquito traps are placed near 43 small ponds and a count of Culex species in the traps is obtained during a baseline period. At the end of the baseline period a spraying program is conducted. One week after the spraying, the traps are cleaned and a second collection period is initiated. Based on an estimate of the size of each pond, the raw counts are converted to a density of Culex mosquitoes per square foot of pond. The difference between the before and after densities is found to have a sample mean of 9.1 and a sample standard deviation of 12.0. Test whether or not the data indicates that the spraying was effective in reducing the density of Culex species mosquitoes.

Category: Hypothesis testing: Inference about two means with paired (dependent) samples

Most Likely Null Hypothesis: The spraying has no effect.

Example 7

Conclusion: Reject the null hypothesis. The chances of getting a mean difference this high are less than 5% with the given sample size.

Example 7

Conclusion: Reject the null hypothesis. The chances of getting a mean difference this high are less than 5% with the given sample size.

Interpretation: The spraying is effective.

Example 8

The mayor of New Orleans guesses that 80% of displaced families plan to return to the area they lived in prior to the storm. In a survey of 150 households relocated to temporary housing after hurricane Katrina, 118 indicate that they intend to return to the places they were living before the storm. Does this sample data support the mayor's estimate.

Example 8

The mayor of New Orleans guesses that 80% of displaced families plan to return to the area they lived in prior to the storm. In a survey of 150 households relocated to temporary housing after hurricane Katrina, 118 indicate that they intend to return to the places they were living before the storm. Does this sample data support the mayor's estimate.

Category: Hypothesis testing about a proportion

Most Likely Null Hypothesis: The proportion intending to return is 80%.

Example 8

Conclusion: Accept the null hypothesis. The sample proportion is in the range you would expect it to be in 95% of samples.

Example 8

Conclusion: Accept the null hypothesis. The sample proportion is in the range you would expect it to be in 95% of samples.

Interpretation: There is no reason to doubt the mayor's estimate of 80%.