Introduction to Geostatistics

8. Formal testing. One-sample tests; two-sample tests; difference in means; difference in proportions. p-values, significance, Type-I errors. One-sided and two-sided tests.

Edzer J. Pebesma

edzer.pebesma@uni-muenster.de
Institute for Geoinformatics (**ifgi**)
University of Münster

summer semester 2007/8,
June 9, 2009

Field work

One-day measurement campaign, in couples of two, to address

- device a research question
- planning of a sampling scheme
- doing the sampling
- entering samples in the computer
- statistical analysis (graphs, confidence intervals, tests)
- (brief) reporting

Field work

One-day measurement campaign, in couples of two, to address

- device a research question
- planning of a sampling scheme
- doing the sampling
- entering samples in the computer
- statistical analysis (graphs, confidence intervals, tests)
- (brief) reporting

Field work

One-day measurement campaign, in couples of two, to address

- device a research question
- planning of a sampling scheme
- doing the sampling
- entering samples in the computer
- statistical analysis (graphs, confidence intervals, tests)
- (brief) reporting

Field work

One-day measurement campaign, in couples of two, to address

- device a research question
- planning of a sampling scheme
- doing the sampling
- entering samples in the computer
- statistical analysis (graphs, confidence intervals, tests)
- (brief) reporting

Field work

One-day measurement campaign, in couples of two, to address

- device a research question
- planning of a sampling scheme
- doing the sampling
- entering samples in the computer
- statistical analysis (graphs, confidence intervals, tests)
- (brief) reporting

Field work

One-day measurement campaign, in couples of two, to address

- device a research question
- planning of a sampling scheme
- doing the sampling
- entering samples in the computer
- statistical analysis (graphs, confidence intervals, tests)
- (brief) reporting

Hypothesis testing

Suppose we have the two-sample example, and ask if in the population group A has a mean that differs significantly from that of group B. The approach we've seen last week is to form a confidence interval for the difference $\mu_A - \mu_B$, and check if this overlaps zero. If not, then the means differ significantly.
Given two random samples, $\bar{X}_A$ and $\bar{X}_B$ will always differ, but the difference can be due to

- if $\mu_A = \mu_B$: chance (random sampling),
- if $\mu_A \neq \mu_B$: difference in population means + chance

Hypothesis testing

Suppose we have the two-sample example, and ask if in the population group A has a mean that differs significantly from that of group B. The approach we've seen last week is to form a confidence interval for the difference $\mu_A - \mu_B$, and check if this overlaps zero. If not, then the means differ significantly.

Given two random samples, $\bar{X}_A$ and $\bar{X}_B$ will always differ, but the difference can be due to

- if $\mu_A = \mu_B$: chance (random sampling),
- if $\mu_A \neq \mu_B$: difference in population means + chance

A formal testing procedure

1. Hypotheses: formulate H_0 and H_A
2. Sample size
3. Significance level
4. Sampling distribution of test statistics (Prüfgröße)
5. Critical region
6. Test statistic
7. Conclusion

A formal testing procedure

1. Hypotheses: formulate H_0 and H_A
2. Sample size
3. Significance level
4. Sampling distribution of test statistics (Prüfgröße)
5. Critical region
6. Test statistic
7. Conclusion

A formal testing procedure

1. Hypotheses: formulate H_0 and H_A
2. Sample size
3. Significance level
4. Sampling distribution of test statistics (Prüfgröße)
5. Critical region
6. Test statistic
7. Conclusion

A formal testing procedure

1. Hypotheses: formulate H_0 and H_A
2. Sample size
3. Significance level
4. Sampling distribution of test statistics (Prüfgröße)
5. Critical region
6. Test statistic
7. Conclusion

ifgi

A formal testing procedure

1. Hypotheses: formulate H_0 and H_A
2. Sample size
3. Significance level
4. Sampling distribution of test statistics (Prüfgröße)
5. Critical region
6. Test statistic
7. Conclusion

A formal testing procedure

1. Hypotheses: formulate H_0 and H_A
2. Sample size
3. Significance level
4. Sampling distribution of test statistics (Prüfgröße)
5. Critical region
6. Test statistic
7. Conclusion

A formal testing procedure

1. Hypotheses: formulate H_0 and H_A
2. Sample size
3. Significance level
4. Sampling distribution of test statistics (Prüfgröße)
5. Critical region
6. Test statistic
7. Conclusion

One-sample test

For example, for the `students` Length data, test whether the population mean might be 175 cm.

1. $H_0 : \mu = 175$, $H_A : \mu \neq 175$
2. $n = 86$
3. $\alpha = 0.05$
4. Sampling distribution of test statistics: t-distribution with $n - 1 = 85$ degrees of freedom
5. Critical region: from $t_{0.025,85} = -1.99$, to $t_{0.975,85} = 1.99$, so any t outside $[-1.99, 1.99]$ leads to rejecting H_0
6. $t = (\bar{X} - \mu)/\text{SE} = 2.61198$
7. Conclusion: t is in the critical region, so we can reject H_0

Meaning that the sample mean is significantly different from the hypothesized value.
Significant: meaningful, not a result from chance

One-sample test

For example, for the `students` Length data, test whether the population mean might be 175 cm.

1. $H_0 : \mu = 175$, $H_A : \mu \neq 175$
2. $n = 86$
3. $\alpha = 0.05$
4. Sampling distribution of test statistics: t-distribution with $n - 1 = 85$ degrees of freedom
5. Critical region: from $t_{0.025,85} = -1.99$, to $t_{0.975,85} = 1.99$, so any t outside $[-1.99, 1.99]$ leads to rejecting H_0
6. $t = (\bar{X} - \mu)/\text{SE} = 2.61198$
7. Conclusion: t is in the critical region, so we can reject H_0

Meaning that the sample mean is significantly different from the hypothesized value.
Significant: meaningful, not a result from chance

One-sample test

For example, for the `students` Length data, test whether the population mean might be 175 cm.

1. $H_0 : \mu = 175$, $H_A : \mu \neq 175$
2. $n = 86$
3. $\alpha = 0.05$
4. Sampling distribution of test statistics: t-distribution with $n - 1 = 85$ degrees of freedom
5. Critical region: from $t_{0.025,85} = -1.99$, to $t_{0.975,85} = 1.99$, so any t outside $[-1.99, 1.99]$ leads to rejecting H_0
6. $t = (\bar{X} - \mu)/\mathrm{SE} = 2.61198$
7. Conclusion: t is in the critical region, so we can reject H_0

Meaning that the sample mean is significantly different from the hypothesized value.
Significant: meaningful, not a result from chance

One-sample test

For example, for the `students` Length data, test whether the population mean might be 175 cm.

1. $H_0 : \mu = 175$, $H_A : \mu \neq 175$
2. $n = 86$
3. $\alpha = 0.05$
4. Sampling distribution of test statistics: t-distribution with $n - 1 = 85$ degrees of freedom
5. Critical region: from $t_{0.025,85} = -1.99$, to $t_{0.975,85} = 1.99$, so any t outside $[-1.99, 1.99]$ leads to rejecting H_0
6. $t = (\bar{X} - \mu)/\mathrm{SE} = 2.61198$
7. Conclusion: t is in the critical region, so we can reject H_0

Meaning that the sample mean is significantly different from the hypothesized value.
Significant: meaningful, not a result from chance

One-sample test

For example, for the `students` Length data, test whether the population mean might be 175 cm.

1. $H_0 : \mu = 175$, $H_A : \mu \neq 175$
2. $n = 86$
3. $\alpha = 0.05$
4. Sampling distribution of test statistics: t-distribution with $n - 1 = 85$ degrees of freedom
5. Critical region: from $t_{0.025,85} = -1.99$, to $t_{0.975,85} = 1.99$, so any t outside $[-1.99, 1.99]$ leads to rejecting H_0
6. $t = (\bar{X} - \mu)/\text{SE} = 2.61198$
7. Conclusion: t is in the critical region, so we can reject H_0

Meaning that the sample mean is significantly different from the hypothesized value.
Significant: meaningful, not a result from chance

One-sample test

For example, for the `students` Length data, test whether the population mean might be 175 cm.

1. $H_0 : \mu = 175$, $H_A : \mu \neq 175$
2. $n = 86$
3. $\alpha = 0.05$
4. Sampling distribution of test statistics: t-distribution with $n - 1 = 85$ degrees of freedom
5. Critical region: from $t_{0.025,85} = -1.99$, to $t_{0.975,85} = 1.99$, so any t outside $[-1.99, 1.99]$ leads to rejecting H_0
6. $t = (\bar{X} - \mu)/\mathrm{SE} = 2.61198$
7. Conclusion: t is in the critical region, so we can reject H_0

Meaning that the sample mean is significantly different from the hypothesized value.
Significant: meaningful, not a result from chance

One-sample test

For example, for the `students` Length data, test whether the population mean might be 175 cm.

1. $H_0 : \mu = 175$, $H_A : \mu \neq 175$
2. $n = 86$
3. $\alpha = 0.05$
4. Sampling distribution of test statistics: t-distribution with $n - 1 = 85$ degrees of freedom
5. Critical region: from $t_{0.025,85} = -1.99$, to $t_{0.975,85} = 1.99$, so any t outside $[-1.99, 1.99]$ leads to rejecting H_0
6. $t = (\bar{X} - \mu)/\text{SE} = 2.61198$
7. Conclusion: t is in the critical region, so we can reject H_0

Meaning that the sample mean is significantly different from the hypothesized value.
Significant: meaningful, not a result from chance

Testing by using confidence intervals

As seen in the previous lecture, the 95% confidence interval for the sample mean is

$$[175.7803, 180.7546]$$

If H_0 does not lie in the central 95% confidence interval, we can reject it.
Note the following

- confidence intervals are on the scale of $\bar{X}$ and μ, test values always on the scale of t, z, etc
- confidence intervals immediately show all the H_0 that would be rejected, and those that would not
- steps 5 and 6 are different: whereas the CI approach uses the critical t and SE to find the boundaries to compare H_0 against, formal tests compare the t test statistic against a critical t value.

Testing by using confidence intervals

As seen in the previous lecture, the 95% confidence interval for the sample mean is

$$[175.7803, 180.7546]$$

If H_0 does not lie in the central 95% confidence interval, we can reject it.
Note the following

- confidence intervals are on the scale of $\bar{X}$ and μ, test values always on the scale of t, z, etc
- confidence intervals immediately show all the H_0 that would be rejected, and those that would not
- steps 5 and 6 are different: whereas the CI approach uses the critical t and SE to find the boundaries to compare H_0 against, formal tests compare the t test statistic against a critical t value.

Testing by using confidence intervals

As seen in the previous lecture, the 95% confidence interval for the sample mean is

$$[175.7803, 180.7546]$$

If H_0 does not lie in the central 95% confidence interval, we can reject it.

Note the following

- confidence intervals are on the scale of $\bar{X}$ and μ, test values always on the scale of t, z, etc
- confidence intervals immediately show all the H_0 that would be rejected, and those that would not
- steps 5 and 6 are different: whereas the CI approach uses the critical t and SE to find the boundaries to compare H_0 against, formal tests compare the t test statistic against a critical t value.

By computer:

```
> load("students.RData")
> attach(students)
> t.test(Length, mu = 175)

        One Sample t-test

data:  Length
t = 3.3814, df = 148, p-value = 0.0009227
alternative hypothesis: true mean is not equal to 175
95 percent confidence interval:
 176.2607 179.8064
sample estimates:
mean of x
 178.0336
```

Where is α?

Statistics programs (such as R) do not ask for an α, but rather give a p-value. This is the probability of wrongly rejecting H_0. If p-value $< \alpha$, you reject H_0, else you do not reject H_0.

About not rejecting H_0

Not rejecting H_0 does never mean that H_0 is true, but merely that it is not in conflict with the data. As the confidence interval shows, there is a large collection of H_0 hypotheses possible, i.e. in agreement with the data, so claiming there is one that is true is quite opportunistic. Furthermore, a so-called *point-hypothesis* sucha as H_0: $\mu = 175$ is quite unlikely to ever be true, as it means $\mu = 175.000000000...$

About not rejecting H_0

Not rejecting H_0 does never mean that H_0 is true, but merely that it is not in conflict with the data. As the confidence interval shows, there is a large collection of H_0 hypotheses possible, i.e. in agreement with the data, so claiming there is one that is true is quite opportunistic. Furthermore, a so-called *point-hypothesis* sucha as H_0: $\mu = 175$ is quite unlikely to ever be true, as it means $\mu = 175.000000000...$

About not rejecting H_0

Not rejecting H_0 does never mean that H_0 is true, but merely that it is not in conflict with the data. As the confidence interval shows, there is a large collection of H_0 hypotheses possible, i.e. in agreement with the data, so claiming there is one that is true is quite opportunistic. Furthermore, a so-called *point-hypothesis* sucha as H_0: $\mu = 175$ is quite unlikely to ever be true, as it means $\mu = 175.000000000...$

Two-sample tests

E.g. difference in means:

- Step 1: $H_0 : \mu_1 = \mu_2$, and
- Step 4: $t = \frac{\bar{X}_1 - \bar{X}_2}{\mathrm{SE}}$ follows a t distribution
- SE: see confidence intervals
- ... with the assumption that $\sigma_1^2 = \sigma_2^2$

Two-sample tests

E.g. difference in means:

- Step 1: $H_0 : \mu_1 = \mu_2$, and
- Step 4: $t = \frac{\bar{X}_1 - \bar{X}_2}{\text{SE}}$ follows a t distribution
- SE: see confidence intervals
- ... with the assumption that $\sigma_1^2 = \sigma_2^2$

Two-sample tests

E.g. difference in means:

- Step 1: $H_0 : \mu_1 = \mu_2$, and
- Step 4: $t = \frac{\bar{X}_1 - \bar{X}_2}{\text{SE}}$ follows a t distribution
- SE: see confidence intervals
- ... with the assumption that $\sigma_1^2 = \sigma_2^2$

Two-sample tests

E.g. difference in means:

- Step 1: $H_0 : \mu_1 = \mu_2$, and
- Step 4: $t = \frac{\bar{X}_1 - \bar{X}_2}{\mathrm{SE}}$ follows a t distribution
- SE: see confidence intervals
- ... with the assumption that $\sigma_1^2 = \sigma_2^2$

Type I and Type II errors

Of course we take a risk to wrongly rejecting a true H_0, of α. There's however also a risk that we wrongly *not* reject a false H_0, which is called β.

Test result	Truth	
	H_0 true	H_0 false
Reject H_0	Type I error, α	OK, (1-β)
Do not reject H_0	OK (1-α)	Type II error, β

β can be controlled by n, and is smaller for larger n. You can compute β under a given H_A (WW: 302-307; next week more on this)

One-sided vs. two-sided tests

Usually the H_A is a simple denial of H_0, as in
H_0: $\mu_1 = \mu_2$
H_A: $\mu_1 \neq \mu_2$ (implying $\mu_1 < \mu_2$ or $\mu_1 > \mu_2$)
We might however be interested in only one type of alternative, e.g.
H_A: $\mu_1 < \mu_2$
In that latter case, as $t = \frac{\bar{X}_1 - \bar{X}_2}{SE}$ we can take the critical region as only the negative t values, and ignore the positive ones. The critical region then is then anything below $t_{0.05, n_1+n_2-2}$
Compare this with one-sided confidence intervals.

```
> t.test(Length, mu = 175, alternative = "less")

        One Sample t-test

data:  Length
t = 3.3814, df = 148, p-value = 0.9995
alternative hypothesis: true mean is less than 175
95 percent confidence interval:
     -Inf 179.5185
sample estimates:
mean of x
 178.0336

> t.test(Length, mu = 175, alternative = "greater")

        One Sample t-test

data:  Length
t = 3.3814, df = 148, p-value = 0.0004614
alternative hypothesis: true mean is greater than 175
95 percent confidence interval:
 176.5486      Inf
sample estimates:
mean of x
 178.0336
```