7. Significance tests for linear regression

- reviews on hypothesis testing
- regression coefficient test

Hypothesis tests

elements of statistical tests

- null hypothesis, alternative hypothesis
- test statistics
- rejection region
- type of errors: type I and type II errors
- confidence intervals, p-values

examples of hypothesis tests:

- hypothesis tests for the mean, and for comparing the means
- hypothesis tests for the variance, and for comparing variances

Testing procedures

a test consists of

- providing a statement of the hypotheses (H_0 (null) and H_1 (alternative))
- giving a rule that dictates if H_0 should be rejected or not

the decision rule involves a test statistic calculated on observed data

the Neyman-Pearson methodology partitions the sample space into two regions

the set of values of the test statistic for which:

the null hypothesis is rejected	**rejection region**
we fail to reject the null hypothesis	**acceptance region**

Test errors

since a test statistic is random, the same test can lead to different conclusions

- **type I error**: the test leads to *reject* H_0 when it is *true*
- **type II error:** the test *fails* to reject H_0 when it is *false*; sometimes called false alarm

probabilities of the errors:

- let β be the probability of type II error
- the **size** of a test is the probability of a type I error and denoted by α
- the **power** of a test is the probability of rejecting a false H_0 or $(1 - \beta)$

α is known as **significance level** and typically controlled by an analyst

for a given α, we would like β to be as small as possible

Some common tests

- normal test
- t-test
- F-test
- Chi-square test

e.g. a test is called a t-test if the test statistic follows t-distribution

two approaches of hypothesis test

- critical value approach
- p-value approach

Critical value approach

Definition: the critical value (associated with a significance level α) is the value of the known distribution of the test statistic such that the probability of type I error is α

steps involved this test

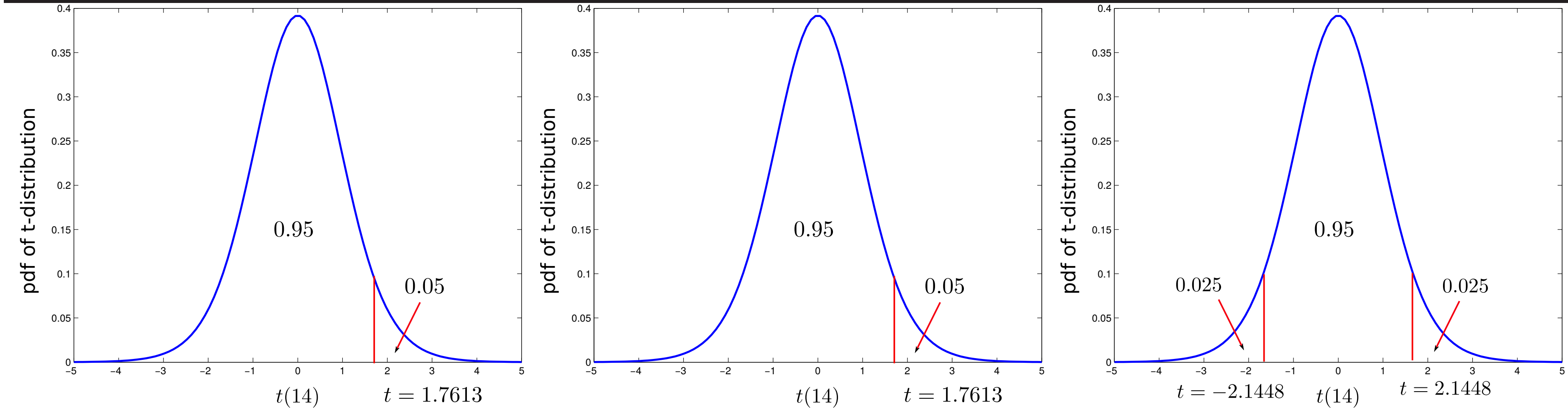
1. define the null and alternative hypotheses.
2. assume the null hypothesis is true and calculate the value of the test statistic
3. set a small significance level (typically $\alpha = 0.01, 0.05$, or 0.10) and determine the corresponding critical value
4. compare the test statistic to the critical value

condition	decision
the test statistic is more extreme than the critical value	reject H_0
the test statistic is less extreme than the critical value	accept H_0

example: hypothesis test on the population mean

- samples $N = 15$, $\alpha = 0.05$
- the test statistic is $t^* = \frac{\bar{x}-\mu}{s/\sqrt{N}}$ and has t-distribution with $N-1$ df

test	H_0	H_1	critical value	reject H_0 if
right-tail	$\mu = 3$	$\mu > 3$	$t_{\alpha,N-1}$	$t^* \geq t_{\alpha,N-1}$
left-tail	$\mu = 3$	$\mu < 3$	$-t_{\alpha,N-1}$	$t^* \leq -t_{\alpha,N-1}$
two-tail	$\mu = 3$	$\mu \neq 3$	$-t_{\alpha/2,N-1}, t_{\alpha/2,N-1}$	$t^* \geq t_{\alpha/2,N-1}$ or $t^* \leq -t_{\alpha/2,N-1}$



p-value approach

Definition: the p-value is the probability of observing a more extreme test statistic in the direction of H_1 than the one observed, by assuming that H_0 were true

steps involved this test

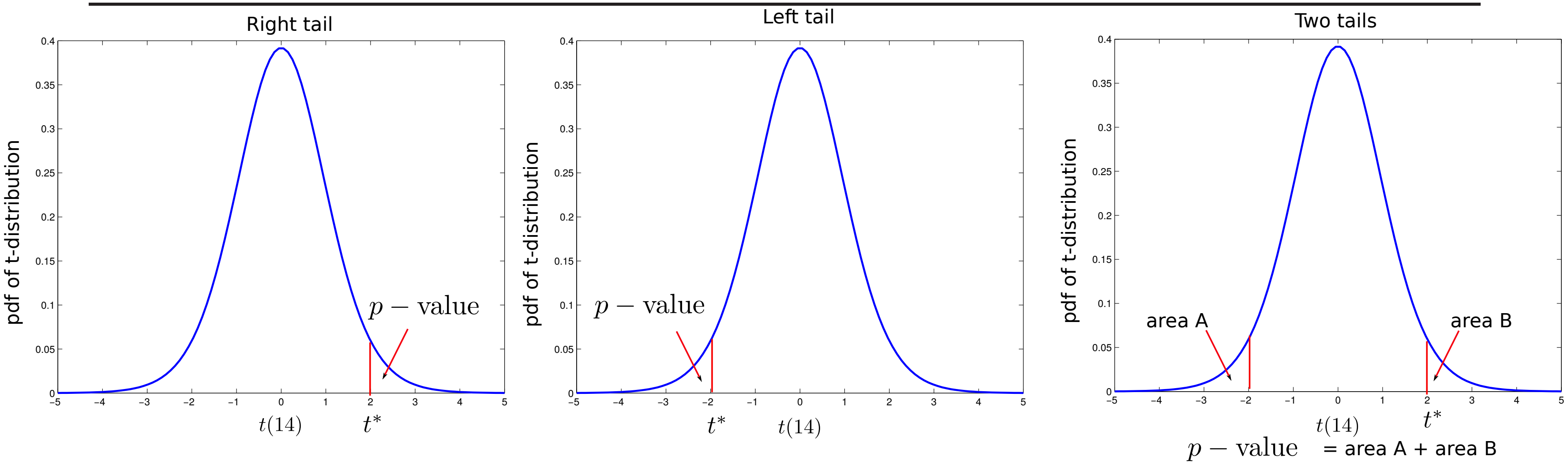
1. define the null and alternative hypotheses.
2. assume the null hypothesis is true and calculate the value of the test statistic
3. calculate the p-value using the known distribution of the test statistic
4. set a significance level α (small value such as $0.01, 0.05$)
5. compare the p-value to α

condition	decision
p-value $\leq \alpha$	reject H_0
p-value $\geq \alpha$	accept H_0

example: hypothesis test on the population mean (same as on page 7-7)

- samples $N = 15$, $\alpha = 0.01$ (have only a 1% chance of making a Type I error)
- suppose the test statistic (calculated from data) is $t^* = 2$

test	H_0	H_1	p-value expression	p-value
right-tail	$\mu = 3$	$\mu > 3$	$P(t_{14} \geq 2)$	0.0127
left-tail	$\mu = 3$	$\mu < 3$	$P(t_{14} \leq -2)$	0.0127
two-tail	$\mu = 3$	$\mu \neq 3$	$P(t_{14} \geq 2) + P(t_{14} \leq -2)$	0.0255



right-tail/left-tail tests: reject H_0, two-tail test: accept H_0

the two approaches assume H_0 were true and determine

p-value	critical value
the probability of observing a more extreme test statistic in the direction of the alternative hypothesis than the one observed	whether or not the observed test statistic is more extreme than would be expected (called critical value)

the null hypothesis is rejected if

p-value	critical value
$p - \text{value} \leq \alpha$	test statistic $\geq$ critical value

Significance tests for linear regression

- reviews on hypothesis testing
- **regression coefficient test**

Recap of linear regression

a linear regression model is

$$y = X\beta + u, \quad X \in \mathbf{R}^{N \times n}$$

homoskedasticity assumption: u_i has the same variance for all i, given by σ^2

- prediction (fitted) error: $\hat{u} := \hat{y} - y = X\hat{\beta} - y$
- residual sum of squares: $\mathrm{RSS} = \|\hat{u}\|_2^2$
- a consistent estimate of σ^2: $s^2 = \mathrm{RSS}/(N-n)$
- $(N-n)s^2 \sim \chi^2(N-n)$
- square root of s^2 is called **standard error of the regression**
- $\mathbf{Avar}(\hat{\beta}) = s^2(X^TX)^{-1}$ (estimated asymptotic covariance)

Common tests for linear regression

- testing a hypothesis about a coefficient

$$H_0 : \beta_k = 0 \quad \text{VS} \quad H_1 : \beta_k \neq 0$$

  we can use both t and F statistics

- testing using the fit of the regression

$$H_0 : \text{reduced model} \quad \text{VS} \quad H_1 : \text{full model}$$

  if H_0 were true, the reduced model ($\beta_k = 0$) would lead to smaller prediction error than that of the full model ($\beta_k \neq 0$)

Testing a hypothesis about a coefficient

statistics for testing hypotheses:

$$H_0 : \beta_k = 0 \quad \text{VS} \quad H_1 : \beta_k \neq 0$$

- $\dfrac{\hat{\beta}_k}{\sqrt{s^2((X^T X)^{-1})_{kk}}} \sim t_{N-n}$
- $\dfrac{(\hat{\beta}_k)^2}{\sqrt{s^2((X^T X)^{-1})_{kk}}} \sim F_{1,N-n}$

the above statistics are Wald statistics (see derivations in Greene book)

- the term $\sqrt{s^2((X^T X)^{-1})_{kk}}$ is referred to **standard error of the coefficient**
- the expression of SE can be simplified or derived in many ways (please check)
- e.g. R use t-statistic (two-tail test)

Testing using the fit of the regression

hypotheses are based on the fitting quality of reduced/full models

$$H_0 : \text{reduced model} \quad \text{VS} \quad H_1 : \text{full model}$$

reduced model: $\beta_k = 0$ and full model: $\beta_k \neq 0$

the F-statistic used in this test

$$\frac{(\mathrm{RSS}_R - \mathrm{RSS}_F)}{\mathrm{RSS}_F/(N-n)} \sim F(1, N-n)$$

- RSS_R and RSS_F are the residual sum squares of reduced and full models
- RSS_R cannot be smaller than RSS_F, so if H_0 were true, then the F statistic would be zero
- e.g. `fitlm` in MATLAB use this F statistic, or in ANOVA table

MATLAB example

perform t-test using $\alpha = 0.05$ and the true parameter is $\beta = (1, 0, -1, 0.5)$

realization 1: $N = 100$

```
>> [btrue b SE pvalue2side]  =
    1.0000    1.0172    0.1087    0.0000
         0    0.1675    0.0906    0.0675
   -1.0000   -1.0701    0.1046    0.0000
    0.5000    0.5328    0.1007    0.0000
```

- $\hat{\beta}$ is close to β
- it's not clear if $\hat{\beta}_2$ is zero but the test decides $\hat{\beta}_2 = 0$
- note that all coefficients have pretty much the same SE

realization 2: $N = 10$

```
>> [btrue b SE pvalue2side]  =
    1.0000    1.0077    0.2894    0.0131
         0    0.1282    0.4342    0.7778
   -1.0000   -1.5866    0.2989    0.0018
    0.5000    0.2145    0.2402    0.4062
```

realization 3: $N = 10$

```
>> [btrue b SE pvalue2side]  =
    1.0000    0.8008    0.3743    0.0762
         0   -0.5641    0.5442    0.3399
   -1.0000   -1.1915    0.5117    0.0588
    0.5000    0.6932    0.4985    0.2137
```

- some of $\hat{\beta}$ is close to the true value but some is not
- the test 2 decides $\hat{\beta}_2$ and $\hat{\beta}_4$ are zero while the test 3 decides all β are zero
- the sample size N affects type II error (fails to reject H_0) and we get different results from different data sets

Summary

- common tests are available in many statistical softwares, e.g, `minitab`, `lm` in R, `fitlm` in MATLAB,
- one should use with care and interpret results correctly
- an estimator is random; one cannot trust its value calculated based on a data set
- examining statistical properties of an estimator is preferred

References

W.H. Greene, *Econometric Analysis*, Prentice Hall, 2008

Review of Basic Statistics (online course)

`https://onlinecourses.science.psu.edu/statprogram`

Stat 501 (online course)

https://onlinecourses.science.psu.edu/stat501