Stat GR5205 Lecture 4

Jingchen Liu

Department of Statistics
Columbia University

Hypothesis testing

- Testing of hypotheses
- Basic setting
  1. Two families of models without overlap: the null hypothesis (H_0) and the alternative hypothesis (H_1)
  2. Which family were the data sampled from?

Hypothesis testing

- The decision: reject the null or do not reject the null
- Two types of errors
  - Type I error
  - Type II error

Hypothesis testing

- The rationale: assuming the null hypothesis and trying to reach contradiction
- Rejection region
- Test statistic: differentiating the null and the alternative

Hypothesis testing

- Formulation

$$H_0 : \beta_1 = 0 \quad \textit{versus} \quad H_1 : \beta_1 \neq 0$$

- $\hat{\beta}_1 = 0.80$ and $SD(\hat{\beta}_1) = 0.10$.
- The sampling distribution of β_1

Hypothesis testing

- The rationale
- Test statistic
- Reference distribution
- p-value: the probability of observing a test statistic that is as extreme as or more extreme than the observed one.
- The null hypothesis is rejected is the p-value is less than a threshold, such as 5%.

Hypothesis testing – reference distribution

- Reference distribution
- The Z-test
- The t-test
- The p-value

Hypothesis testing

- Size of a test: the probability of making type I error.
- Power of a test: the probability of rejecting the null under the alternative hypothesis

Hypothesis testing

- The Z-test

$$z = \frac{\hat{\beta}_1}{\sigma\sqrt{\frac{1}{\sum(x_i - \bar{x})}}} \sim N(0,1)$$

- The t-test

$$t = \frac{\hat{\beta}_1}{\hat{\sigma}\sqrt{\frac{1}{\sum(x_i - \bar{x})}}} \sim t_{n-2}$$

Hypothesis testing – *p*-value

- Two-sided *p*-value: $H_0 : \beta_1 = 0 \quad H_1 : \beta_1 \neq 0$

$$P(|Z| \geq |z|) \quad or \quad P(|t_{n-2}| \geq |t|)$$

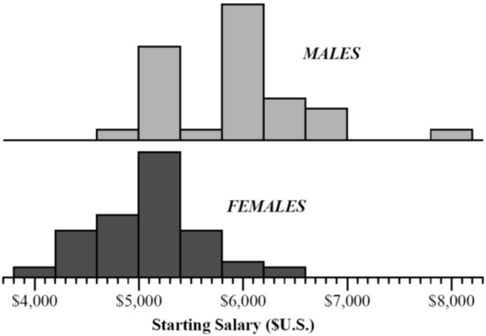
- One-sided *p*-value: $H_0 : \beta_1 \leq 0 \quad H_1 : \beta_1 > 0$

$$P(Z \geq z) \quad or \quad P(t_{n-2} \geq t)$$

One-sample and two-sample t-test

Display 1.4 **p. 5**

Frequency histograms for male and female starting salaries



One-sample and two-sample t-test

- $\bar{Y}_{MALE} = 5957$, $\bar{Y}_{FEMALE} = 5139$
- $\bar{Y}_{MALE} - \bar{Y}_{FEMALE} = 818$, $\hat{\sigma} = 596$
- $$t - statistic = \frac{\bar{Y}_{MALE} - \bar{Y}_{FEMALE}}{\hat{\sigma}\sqrt{\frac{1}{n_2} + \frac{1}{n_2}}} = 6.2926.$$

Hypothesis testing – Duality

- Hypothesis testing and confidence interval
- The null is rejected if the confidence interval does not cover the hypothesized value. Justification!

Analysis of variance

- The decomposition

$$\sum(y_i - \bar{y})^2 = \sum(y_i - \hat{y}_i)^2 + \sum(\hat{y}_i - \bar{y})^2$$

$$SS_{total} = SS_{error} + SS_{regression}$$

- The analysis of variance table

	d.f.	Sum Sq	Mean Sq	F-value	p-value
x	$p-1$	$\sum(\hat{y}_i - \bar{y})^2$	$\frac{\sum(\hat{y}_i-\bar{y})^2}{p-1}$	$\frac{(n-p)\sum(\hat{y}_i-\bar{y})^2}{(p-1)\sum(y_i-\hat{y}_i)^2}$	*
Residuals	$n-p$	$\sum(y_i - \hat{y}_i)^2$	$\frac{\sum(y_i-\hat{y}_i)^2}{n-p}$		
Total	$n-1$	$\sum(y_i - \bar{y})^2$			

Analysis of variance

- The analysis of variance table of the Iris setosa data

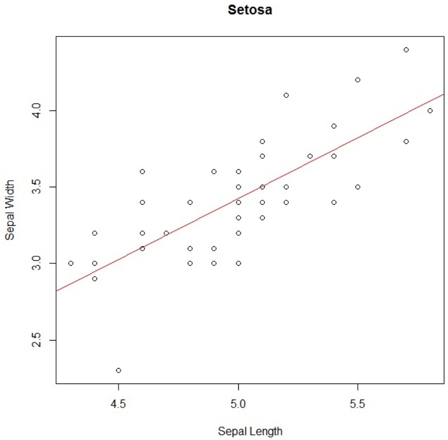
	d.f.	Sum Sq	Mean Sq	F-value	p-value
x	1	3.9	3.2	59.0	7×10^{-10}
Residuals	48	3.2	0.066		
Total	49	7.1			

- F-test and t-test

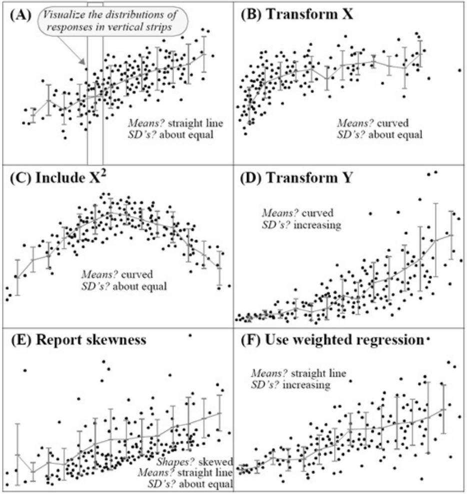
Inferential tools of simple linear model

- The four assumptions
- Point estimate
  - Understanding the least squares estimate
  - variance estimate
  - Frequentist's distribution
- Interval estimate
  - Regression coefficients
  - Prediction: conditional mean, future observation, simultaneous confidence band
- Hypothesis testing
  - Z-test
  - t-test: special case – two sample test
  - F-test: analysis of variance, R^2

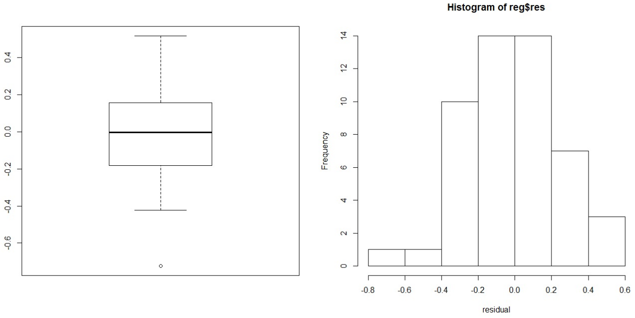
Diagnosis – linearity



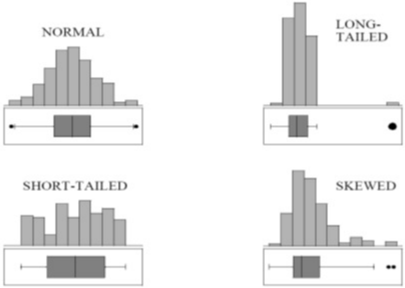
Some possible deviation away from the assumption



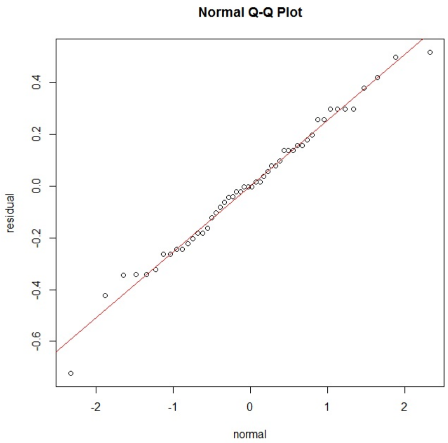
Diagnosis – graphical analysis



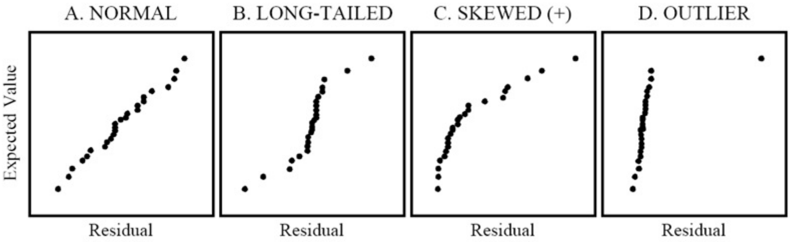
Diagnosis – box plot



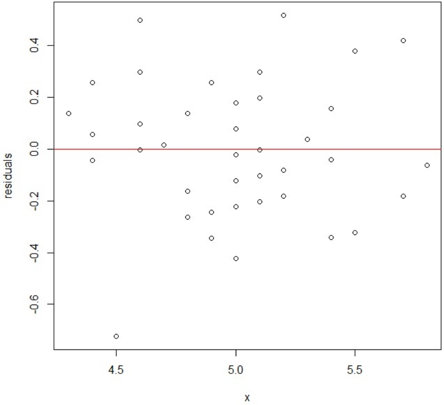
Diagnosis – quantile-quantile plot



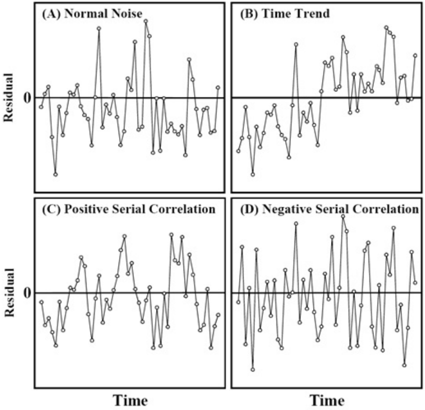
Diagnosis – quantile-quantile plot



Diagnosis – residuals



Some possible deviation away from the assumption



Diagnostic test

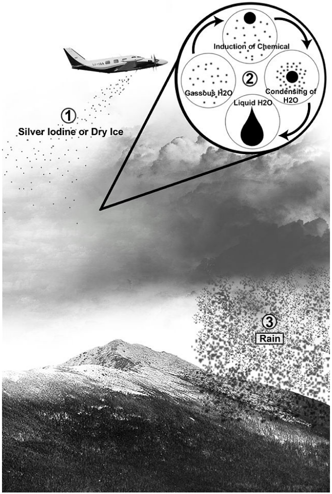
- Pure significant test
- Equal variance test: Brown-Forsythe test and Levene's test

Howard Levene

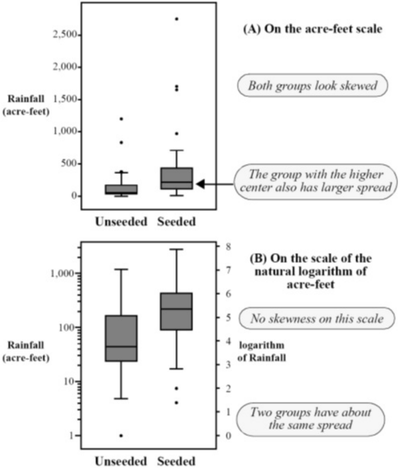
Transformation 101

- Logarithm!

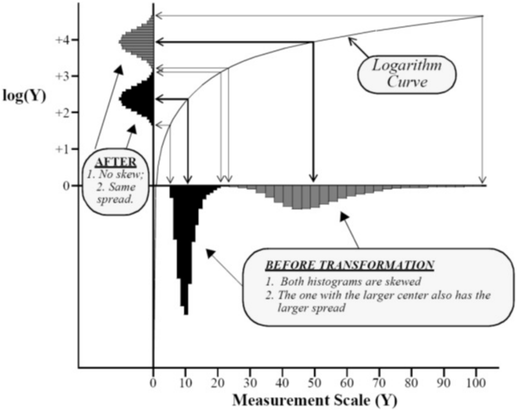
Logarithm


Induction of Chemical
Gassous H2O
Condensing of H2O
Liquid H2O
Silver Iodine or Dry Ice
Rain

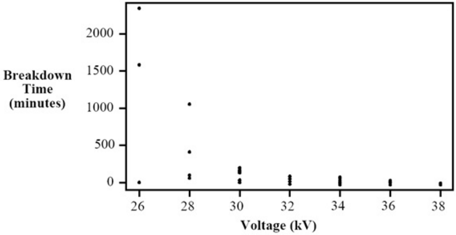
Logarithm


(A) On the acre-feet scale
2,500
2,000
Rainfall (acre-feet)
1,500
1,000
500
0
Both groups look skewed
The group with the higher center also has larger spread
Unseeded
Seeded
(B) On the scale of the natural logarithm of acre-feet
1,000
100
10
1
Rainfall (acre-feet)
8
7
6
5
4
3
2
1
0
No skewness on this scale
logarithm of Rainfall
Two groups have about the same spread
Unseeded
Seeded

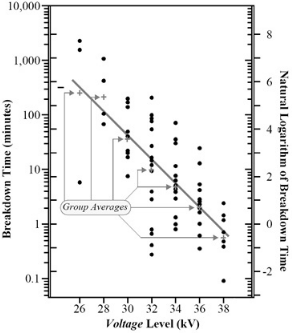
Logarithm


log(Y)
+4
+3
+2
+1
0
Logarithm Curve
AFTER
1. No skew;
2. Same spread.
BEFORE TRANSFORMATION
1. Both histograms are skewed
2. The one with the larger center also has the larger spread
0
10
20
30
40
50
60
70
80
90
100
Measurement Scale (Y)

Logarithm


Breakdown Time (minutes)
2000
1500
1000
500
0
26
28
30
32
34
36
38
Voltage (kV)

Logarithm


10,000
1,000
100
10
1
0.1
Breakdown Time (minutes)
8
6
4
2
0
-2
Natural Logarithm of Breakdown Time
Group Averages
26
28
30
32
34
36
38
Voltage Level (kV)

Box-Cox tranformation

$$f_\lambda(y)$$

- $f_\lambda(y) = (y^\lambda - 1)/\lambda$ if $\lambda \neq 0$; $f_\lambda(y) = \log y$, if $\lambda = 0$.
- Choice of λ: maximum likelihood estimate
- library{car}: box.cox.powers

Some other transformations

$$y \in [0, 1]$$

- Logit transform: $\log \frac{y}{1-y}$
- Probit transform: $F^{-1}(y)$ where $F(x) = P(Z \leq x)$
- etc.