Introduction to Statistics

Lecture 2: Testing and ANOVA

Ta-Wei Huang

Yoctol Info.

September 2, 2016

Table of Contents

Table of Contents

Table of Contents

Table of Contents

One-sample Test for Normal Population

Problem Setting

Let $(X_1, \cdots, X_n)$ be a random sample from $\mathcal{N}(\mu, \sigma^2)$, where μ and σ^2 are unknown parameters.

- Best estimator for μ: $\overline{X} = \frac{X_1 + \cdots + X_n}{n}$.
- Best estimator for σ^2: $\frac{\sum_{i=1}^n \left(X_i - \overline{X}\right)^2}{n-1}$.

Our goal is to compare (or infer)

(1) μ and a constant μ_0 and (2) σ^2 and a constant σ_0^2.

One-sample t Test for Normal Population

Test Setting: One-sample t Test

- Two-tail: $H_0 : \mu = \mu_0$ versus $H_1 : \mu \neq \mu_0$
- Upper-tail: $H_0 : \mu \leq \mu_0$ versus $H_1 : \mu > \mu_0$
- Lower-tail: $H_0 : \mu \geq \mu_0$ versus $H_1 : \mu < \mu_0$

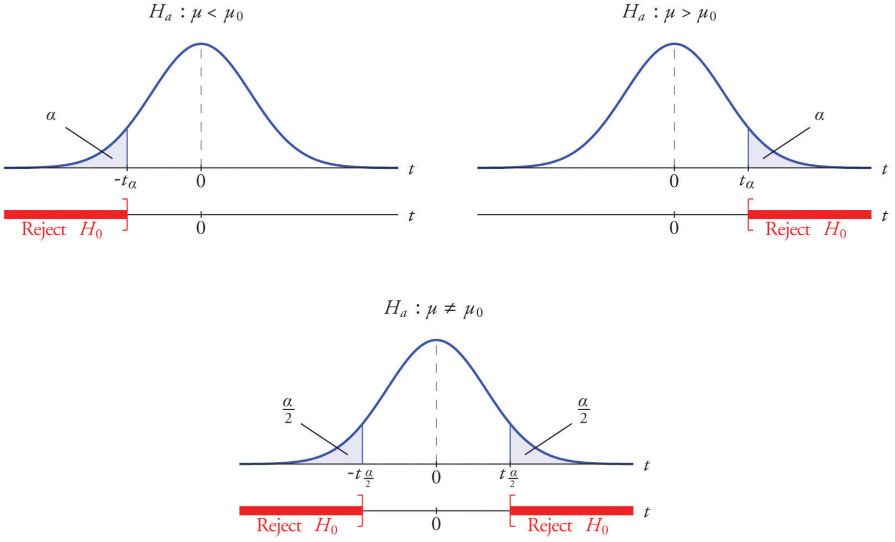
Test Statistic: One-sample t Test

Under $H_0 : \mu = \mu_0$ (we use this setting even for one-tail testing),

$$T = \frac{\overline{X} - \mu}{S/\sqrt{n}} \sim t(n-1),$$

the t distribution with degree of freedom $n-1$.

Rejection Region for One-sample t Test



Example for One-sample t Test 1

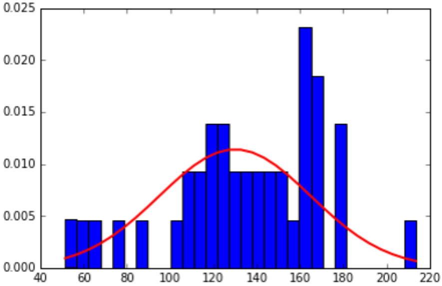
ZARA is considering whether it should release a new pant. One business analyst has calculated that they should sold out at least 125 clothes for one store per month to achieve the break-even point. She randomly selected 40 identical stores to sell the pant for one month. The data has been collected. Could you help ZARA to determine whether to sell the pant?

Example for One-sample t Test 2

Solution to the Result

$H_0 : \mu \leq \mu_0$ versus $H_1 : \mu > \mu_0$, where $\mu_0 = 120$

1. Test statistic: $T = 5.13$, p-value: $0.0000 < \alpha = 0.05$
2. Reject H_0, that is, $\mu > 120$ significantly under $\alpha = 0.05$.



One-sample Test for Normal Variance

Test Setting: One-sample Variance Test

- Two-tail: $H_0 : \sigma^2 = \sigma_0^2$ versus $H_1 : \sigma^2 \neq \sigma_0^2$
- Upper-tail: $H_0 : \sigma^2 \leq \sigma_0^2$ versus $H_1 : \sigma^2 > \sigma_0^2$
- Lower-tail: $H_0 : \sigma^2 \geq \sigma_0^2$ versus $H_1 : \sigma^2 < \sigma_0^2$

Test Statistic: One-sample F Test

Under $H_0 : \sigma^2 = \sigma_0^2$ (we use this setting even for one-tail testing),

$$Q = \frac{(n-1)S^2}{\sigma_0^2} \sim \chi^2_{n-1},$$

the χ^2 distribution with degree of freedom $n-1$.

Example for One-sample χ^2 Test 1

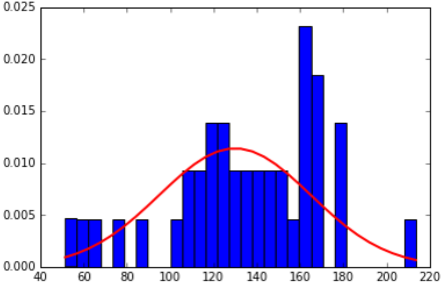
The business analyst thinks that if $\sigma \geq 20$, the risk of selling that pant will be too high. Could you help her to determine whether the risk level is acceptable?

Example for One-sample χ^2 Test 2

Solution to the Result

$H_0 : \sigma^2 \geq \sigma_0^2 = 400$ versus $H_1 : \sigma^2 < \sigma_0^2 = 400$

1. Test statistic: $T = 2.74, \quad p$-value: $0.0046 < \alpha = 0.01$
2. Reject H_0, that is, $\mu > 120$ significantly under $\alpha = 0.01$.



Independent Two-sample Test

Problem Setting

Let $(X_{11}, \cdots, X_{1n_1})$ be a random sample from $\mathcal{N}(\mu_1, \sigma_1^2)$, and $(X_{21}, \cdots, X_{2n_2})$ a random sample from $\mathcal{N}(\mu_2, \sigma_2^2)$.

- $(X_{11}, \cdots, X_{1n_1})$ and $(X_{21}, \cdots, X_{2n_2})$ are independent.
- (μ_1, σ_1^2) and (μ_2, σ_2^2) are unknown parameters.

Our goal is to compare (or infer)

(1) parameters σ_1^2 and σ_2^2 $\Rightarrow$ Special case for Levene test

(2) parameters μ_1 and μ_2. $\Rightarrow$ Independent t test

Levene Test for Equality of Variance

Test Setting: Levene Test

- Two-tail: $H_0 : \sigma_1^2 = \sigma_2^2$ versus $H_1 : \sigma_1^2 \neq \sigma_2^2$
- Can be generalized to test

  $H_0 : \sigma_1^2 = \sigma_2^2 = \cdots = \sigma_k^2$ versus $H_1 : \sigma_i^2 \neq \sigma_j^2$ for some $i \neq j$

The test statistic is complicated, so I omit here. If you are interested in this topic, check the following website:

http://www.itl.nist.gov/div898/handbook/eda/section3/eda35a.htm

Independent Two-sample t Test: Unequal Variance

Test Setting: Independent Two-sample t Test

- Two-tail: $H_0 : \mu_1 = \mu_2$ versus $H_1 : \mu_1 \neq \mu_2$
- Upper-tail: $H_0 : \mu_1 \leq \mu_2$ versus $H_1 : \mu_1 > \mu_2$
- Lower-tail: $H_0 : \mu_1 \geq \mu_2$ versus $H_1 : \mu_2 < \mu_2$

Test Statistic: Independent Two-sample t Test

If $\sigma_1^2 \neq \sigma_2^2$, under $H_0 : \mu_1 = \mu_2$,

$$T = \frac{\overline{X}_1 - \overline{X}_2}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}} \sim t(d.f.),$$

where $d.f. = \frac{(s_1^2/n_1)+(s_2^2/n_2)}{(s_1^2/n_1)^2/(n_1-1)+(s_2^2/n_2)^2/(n_2-1)}$.

Independent Two-sample t Test: Equal Variance

If the variances of two samples are equal, we use the pooled variance $S_p^2 = \frac{(n_1-1)S_1^2+(n_2-1)S_2^2}{n_1+n_2-1}$.

Test Statistic: Independent Two-sample t Test

If $\sigma_1^2 \neq \sigma_2^2$, under $H_0 : \mu_1 = \mu_2$,

$$T = \frac{\overline{X}_1 - \overline{X}_2}{S_p} \sim t(d.f.),$$

where $d.f. = \frac{(s_1^2/n_1)+(s_2^2/n_2)}{(s_1^2/n_1)^2/(n_1-1)+(s_2^2/n_2)^2/(n_2-1)}$.

Example for Independent Two-sample t Test 1

The business analyst are further considering which type of store could sell the pant better. So she want to test the difference between "urban" stores and "rural" stores. She randomly select 50 homogeneous rural stores and 30 homogeneous urban stores, and sell the pant in each store for one month. The data have been collected.

Example for Independent Two-sample t Test 2

Step 1: Test for Equality of Variance

$H_0 : \sigma_1^2 = \sigma_2^2$ versus $H_1 : \sigma_1^2 \neq \sigma_2^2$

1. Test statistic: $W = 0.37$, p-value: $0.5431 > \alpha = 0.05$
2. Do not reject H_0, i.e., no significant difference under $\alpha = 0.05$

Step 2: Test for Difference

$H_0 : \mu_1 \geq \mu_2$ versus $H_1 : \mu_1 < \mu_2$

1. Test statistic: $T = -5.49$, p-value: $0.0000 < \alpha = 0.05$
2. Reject H_0, i.e., $\mu_1 < \mu_2$ significantly under $\alpha = 0.05$

Paired Sample t Test

Problem Setting

Let $(X_{11}, X_{21}), \cdots, (X_{1n}, X_{2n})$ be the repeated measurements on one property of n individuals. Define $D_i = X_{2i} - X_{1i}$.

Assume $(D_1, \cdots, D_n)$ is a random sample from $\mathcal{N}(\mu_D, \sigma_D^2)$.

From the setting, we can see that it's just a one-sample t test!!!

Example: Paired Sample t Test 1

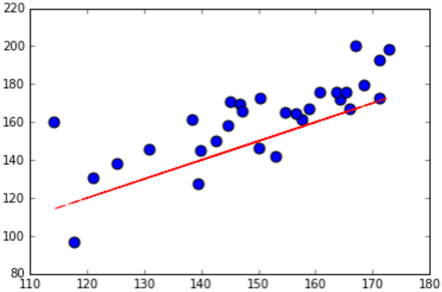
The business analyst decides to sell the pant in rural stores. The feedback is quite good. Now, the business analyst wants to release a marketing campaign around globe. She first chooses 30 stores to test her idea.

Example: Paired Sample t Test 2

Solution and Result

$H_0 : \mu_D \leq 0$ versus $H_1 : \mu_D > 0$.

1. Test statistic: $T = 4.69$, p-value: $0.0001 < \alpha = 0.05$
2. Reject H_0, that is, $\mu_D > 0$ significantly under $\alpha = 0.05$.



Statistical Experimental Design

Experimental design = assign experimental units to treatment conditions.

Spirit of Experimental Design

A good experimental design serves three purposes.

1. Make causal inferences about the relationship between independent variables and a dependent variable.
2. Rule out alternative explanations due to the confounding effects of extraneous variables.
3. Reduce variability within treatment conditions, which makes it easier to detect differences in treatment outcomes.

Statistical Experimental Design

Experimental design = assign experimental units to treatment conditions.

Spirit of Experimental Design

A good experimental design serves three purposes.

1. Make causal inferences about the relationship between independent variables and a dependent variable.
2. Rule out alternative explanations due to the confounding effects of extraneous variables.
3. Reduce variability within treatment conditions, which makes it easier to detect differences in treatment outcomes.

One Factor Experiment - Analysis of Variance

Problem for One-way Fixed Effect Model

Assume the interesting factor separate our model into K groups. In each group i, we collect n_i observations $X_{i,1}, \cdots, X_{i,n}$.

- Model: $X_{i,j} = \alpha + \mu_i + \epsilon_{i,j}$
- Parameters: α - overall mean; μ_i - group mean
- Distribution: $\epsilon_{i,j}$ are iid $\mathcal{N}(0, \sigma^2)$

Our goal is to compare K group means $\mu_1, \mu_2, \cdots, \mu_K$.

One-way ANOVA Overall F Test

Overall F-test

We want to test the (fixed) treatment effect.

$H_0 : \mu_1 = \mu_2 = \cdots = \mu_K$ versus $H_1 : \mu_i \neq \mu_j$ for some (i, j).

We use overall F test to test whether (overall) there exists the treatment effect.

Post-hoc Tukey-Kramer Procedure

Goal: simultaneously to the set of all pairwise comparisons $\mu_i - \mu_j$

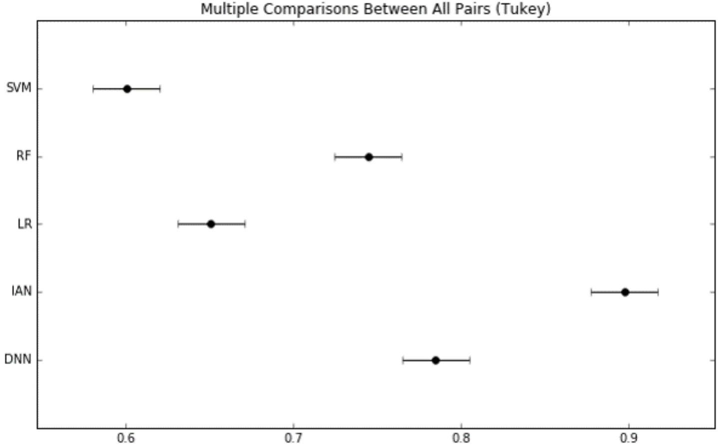
Method: Studentized Range Distribution $Q_{r,v}$

Example: Comparing different classifiers

Ian, a well-known data scientist in Yoctol Info., proposes a new classification model called "deep-ian-say". He uses one dataset, randomly split them into 100 (training, validation) subsets, and then run different algorithms on 100 datasets. Please help him to compare the result.

Example: Comparing different classifiers

Construct the Tukey-Kramer simultaneous confidence interval.



Example: Comparing different classifiers

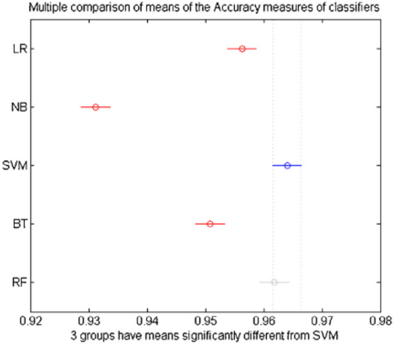
Compare mean difference by Tukey-Kramer testing procedure.

Multiple Comparison of Means - Tukey HSD,FWER=0.05

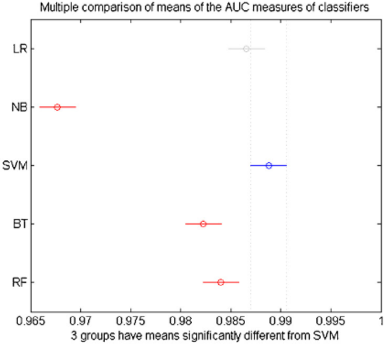
group1	group2	meandiff	lower	upper	reject
DNN	IAN	0.1131	0.0732	0.153	True
DNN	LR	-0.1339	-0.1738	-0.094	True
DNN	RF	-0.0401	-0.08	-0.0002	True
DNN	SVM	-0.1843	-0.2242	-0.1444	True
IAN	LR	-0.2469	-0.2869	-0.207	True
IAN	RF	-0.1532	-0.1931	-0.1133	True
IAN	SVM	-0.2974	-0.3373	-0.2575	True
LR	RF	0.0937	0.0538	0.1336	True
LR	SVM	-0.0504	-0.0903	-0.0105	True
RF	SVM	-0.1442	-0.1841	-0.1043	True

Example: Comparing different classifiers

R. Prashantha, Sumantra Dutta Roya, Pravat K. Mandalb,c, Shantanu Ghoshda. High-Accuracy Detection of Early Parkinson's Disease through Multimodal Features and Machine Learning, International Journal of Medical Informatics.



(a)



(b)