RCDS Statistics II - Imperial College London

Further Hypothesis Testing

Jesús Urtasun Elizari

December 2023

Further hypothesis testing

Chapter 1. Introduction and parameter estimation

i) Parameter estimation

ii) Comparing means. t-test.

iii) Comparing variances. F-test.

Chapter 2. ANOVA and χ^2 test.

i) Standardized variables. z and t distributions

ii) Comparing distributions. χ^2 test.

iii) ANOVA. Comparing more than two groups. F test.

Chapter 3. Multiple hypothesis correction

i) P-values and significance.

ii) Error types in hypothesis testing.

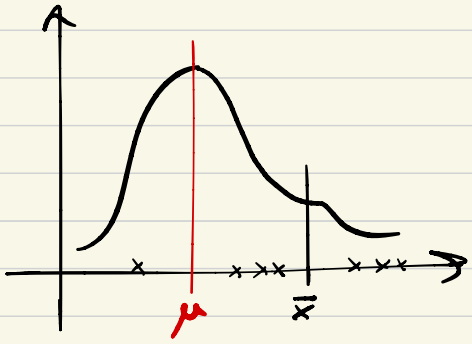
iii) Adjusted p-values. Bonferroni and Benjamin-Hochberg.

1.1 Parameter estimation

We have a set of observations $X = \{x_1, x_2, \dots, x_n\}$

We assume they came from a certain distribution with some true μ and some true σ^2

$f(x)$: $B(x)$, $P(x)$, $Gauss(x)$, ...



Observed average ("sample mean")

$$\bar{x} = \frac{1}{n}(x_1 + x_2 + \dots + x_n) = \frac{1}{n}\sum_{i=1}^{n} x_i$$

Observed variance ("sample variance")

$$s^2 = \frac{1}{n-1}\{(x_1-\bar{x})^2 + (x_2-\bar{x})^2 + \dots + (x_n-\bar{x})^2\}$$

$$= \frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})^2$$

- Sample mean $\bar{x}$ is an estimator of μ
- Sample variance s^2 is an estimator of s^2

* How close are $\bar{x}$, s^2 estimators to the true μ, σ^2 of the population?
* "What was the probability of obtaining a certain sample mean, or a certain sample variance?"

→ Parameter estimation. Further hypothesis testing

Example: $x_1 = \{1, 2, 3\}$ and $x_2 = \{3, 4, 5\}$

Sample mean: $\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i$

Sample variance: $s^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2$

i) Compute sample mean $\bar{x}_1$ and $\bar{x}_2$

$$\bar{x}_1 = \frac{1}{3}(1+2+3) = \frac{6}{3} = 2$$

$$\bar{x}_2 = \frac{1}{3}(3+4+5) = \frac{12}{3} = 4$$

ii) Compute sample variance

$$s_1^2 = \frac{1}{2}\{(1-2)^2 + (2-2)^2 + (3-2)^2\} = 1$$

$$s_2^2 = \frac{1}{2}\{(3-4)^2 + (4-4)^2 + (5-4)^2\} = 1$$

* Hypothesis testing

i) State null and alternative hypothesis H_0, H_1

ii) Collect data / make observations

iii) Compute "statistic" (a number; a function of our data) (*)

iv) Compute p-value (how likely was it to obtain this result)

v) If p-value $< \alpha$, reject H_0. Otherwise, accept H_0

* Different experiments / questions / data are better addressed computing different "statistics" / "statistic tests"

* Comparing means → t-test (Pearson, Gosset, Fisher; 1900s)
* Comparing variances → F-test (Snedecor, Fisher; 1920s)
* Comparing distributions → χ^2 test (Pearson, Fisher; 1920s)

1.2 Comparing means. t-test

Check if two samples have significantly different means

We have two sets of observations $\{x_1\}$ and $\{x_2\}$

of sample size n_1, n_2 and sample means $\bar{x}_1$, $\bar{x}_2$

* t-test is just one kind of statistic (step iii)

useful to compare means between two groups (*)

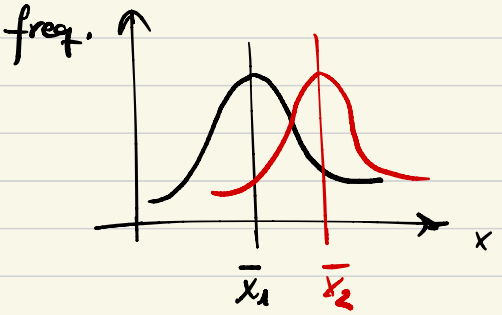
i) Formulate H_0, H_1

H_0 : both samples have the same mean $\bar{x}_1 = \bar{x}_2 = \mu$

H_1 : the samples have different means $\bar{x}_1 \neq \bar{x}_2$

ii) Collect data

Example : University of 2 different types of stars.



$\{x_1\}$: type IV stars ("supergiant")

$\{x_2\}$: type V stars ("hypergiant")

iii) Compute t "statistic" (t variable)

$$\left\{ t = \frac{\bar{x}_1 - \bar{x}_2}{s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \right\} \quad \text{if } \bar{x}_1 = \bar{x}_2 \Rightarrow t = 0 \ \checkmark$$

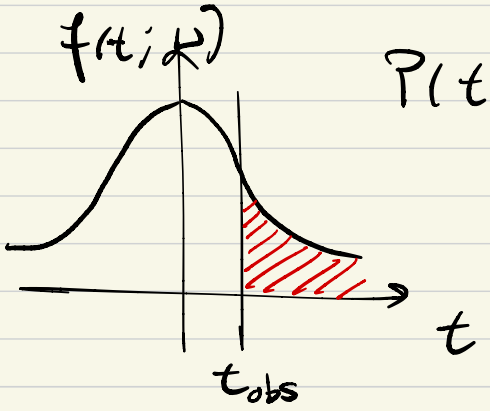
$$s_p = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$$ "pooled" standard deviation.

iv) Compute p-value.

"What was the probability of obtaining a value at least as extreme as the one we got for t."

* If $\bar{x}_1 = \bar{x}_2$, the t variable follows a Student's t distribution with $n_1 + n_2 - 2$ degrees of freedom. (*)

* Assuming H_0 true; what was probability of obtaining t_{obs}?



$$P(t \geq t_{obs}) = 1 - cdf_t(t_{obs}) \quad (*)$$

{ Student's t tables
{ Computer simulations

r) Compare with significance level.

if p-value $> \alpha$; it was likely to obtain this t_{obs}.
We accept H_0. $\bar{x}_1 = \bar{x}_2 = \mu$ ✓

if p-value $< \alpha$; it was unlikely to obtain this t_{obs}.
We reject H_0. $\bar{x}_1$ and $\bar{x}_2$ come from different distributions.

1.3 Comparing variances. F test.

Check if two samples have significantly different variances

We have two sets of observations $\{x_1\}$ and $\{x_2\}$

of sample size n_1, n_2 and sample variances s_1^2, s_2^2

* F-test is just one kind of statistic (step iii)

useful to compare variances between two groups (*)

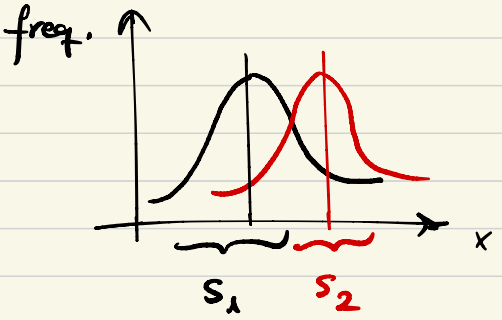
i) Formulate H_0, H_1

H_0 : both samples have same variances; $s_1^2 = s_2^2 = \sigma^2$

H_1 : the samples have different variances; $s_1^2 \neq s_2^2$

ii) Collect data

Example : Luminosity of 2 different types of stars.



$\}x_1\{$: type IV stars ("supergiant")

$\}x_2\{$: type V stars ("hypergiant")

iii) Compute F statistic (F variable)

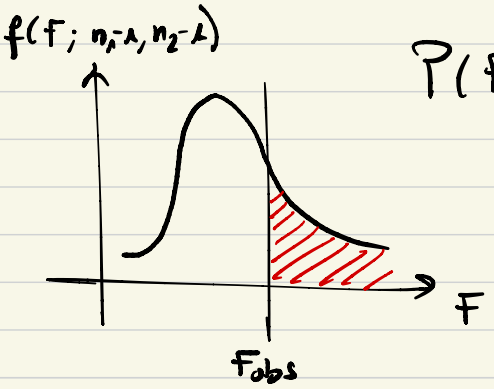
$$F = \frac{s_1^2}{s_2^2} \quad \text{if } s_1^2 = s_2^2 \Rightarrow F = 1 \checkmark$$

iv) Compute p-value.

"What was the probability of obtaining a value at least as extreme as the one we got for F."

* If $s_1^2 = s_2^2$, the F variable follows a F distribution with $(n_1 - 1, n_2 - 2)$ degrees of freedom. (*)

* Assuming H_0 true; what was probability of obtaining F_{obs}?



$$P(F \geq f_{obs}) = 1 - cdf_F(f_{obs}) \quad (*)$$

(*) { F probability tables / Computer simulations }

r) Compare with significance level.

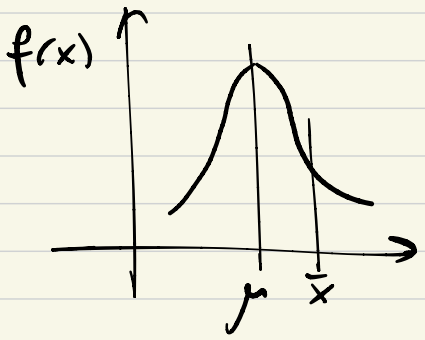
if p-value $> \alpha$; it was likely to obtain this F_{obs}.

We accept H_0. $s_1^2 = s_2^2 = \sigma^2$ ✓

if p-value $< \alpha$; it was unlikely to obtain this t_{obs}.

We reject H_0. s_1^2 and s_2^2 come from different distributions.

2.1 Standardized variables



We have a series of observations $X = \{x_1, x_2, \ldots, x_n\}$

We assume they come from a certain $f(x)$

Sample mean $\bar{x}$ as estimator of true μ

Sample variance s^2 as estimator of true σ^2

* Hypothesis testing. How close are our $\bar{x}$ and s^2 estimators to the true population mean μ and var. σ^2?

* Build a "statistic" (a variable, a quantity)

* Standardized variable

If we know the distribution our samples come from ("percentile")

$$\frac{\bar{x}-\mu}{\sigma} = \uparrow \frac{\bar{x}-\mu}{s/\sqrt{n}} \equiv \begin{cases} z \text{ variable (if } n \text{ large)} \\ t \text{ variable (if } n \text{ small)} \end{cases}$$

↑ Central Limit Theorem

i) If n large ($n \gtrsim 30$)

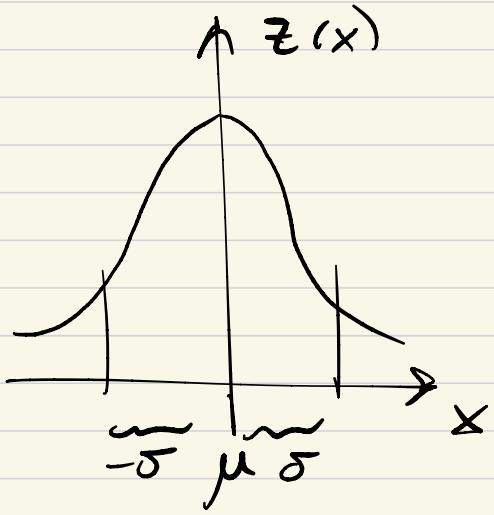
$$z = \frac{\bar{x}-\mu}{\sigma/\sqrt{n}} \text{ follows a Normal distribution}$$

ii) If n small ($n \lesssim 30$)

$$t = \frac{\bar{x}-\mu}{\sigma/\sqrt{n}} \text{ follows a Student's t distribution}$$

i) Normal distribution ("Gaussian") with mean μ and std σ

$$\left\{ z(x;\mu,\sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2\sigma^2}(x-\mu)^2} \right\}$$



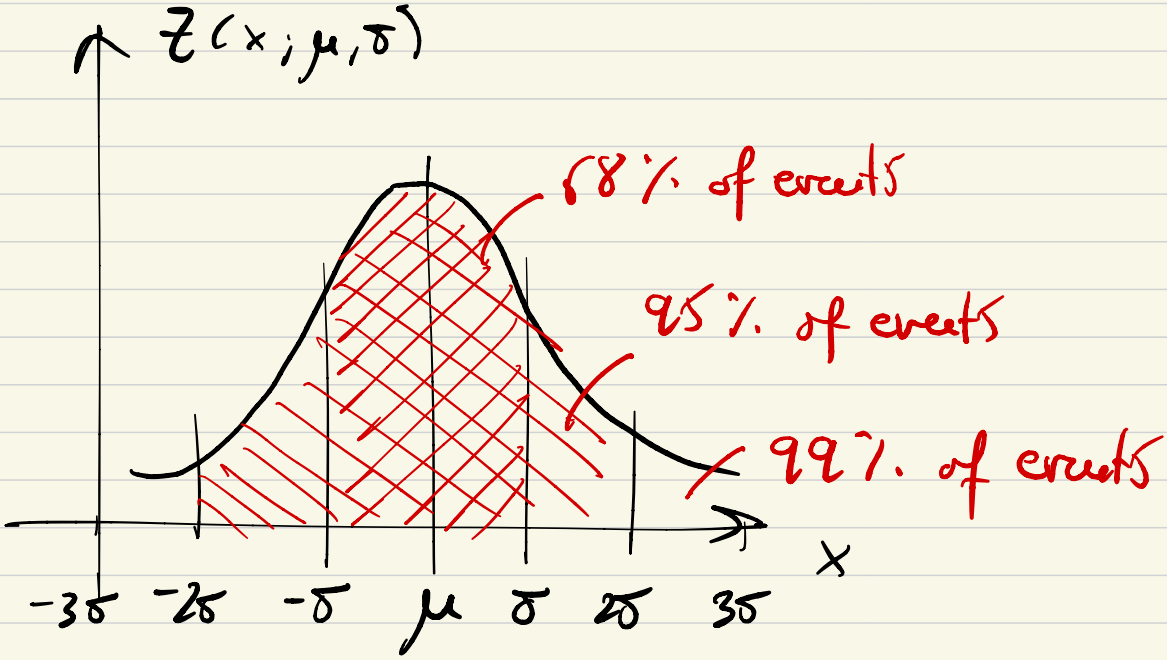
Simplified case $\mu = 0$; $\sigma = 1$

$$\left\{ z(x;0,1) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} \right\}$$

Sometimes written as $f(z;\mu,\sigma)$

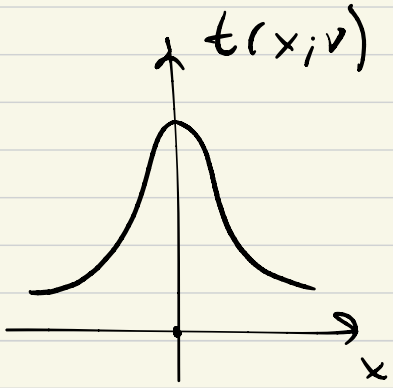
* Confidence intervals in a Normal distribution

1σ	confidence interval	$(\mu-\sigma, \mu+\sigma)$	contains 68 % events
2σ	//	$(\mu-2\sigma, \mu+2\sigma)$	// 95 % events
3σ	//	$(\mu-3\sigma, \mu+3\sigma)$	// 99 % events.



ii) Student's t distribution with ν degrees of freedom

$$\left\{ t(x;\nu) \sim \left(1 + \frac{x^2}{\nu}\right)^{-\frac{1}{2}(\nu+1)} \right\}$$



Simplified case with $\nu = 1$

$$\left\{ t(x;1) \sim (1+x^2)^{-1} = \frac{1}{1+x^2} \right\}$$

Sometimes written as $f(t;\nu)$

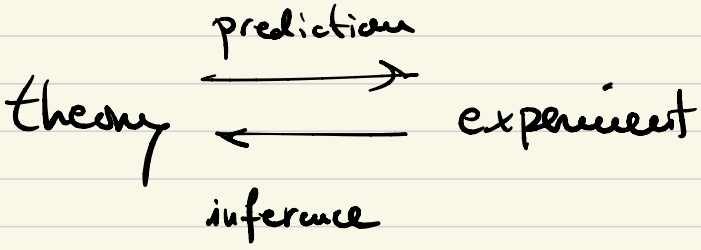
* As the number of degrees of freedom ν increase, the t distribution tends more and more to a Gaussian.

* Confidence intervals in Student's t distribution

1σ	confidence interval	$(\mu \pm \sigma)$
2σ	//	$(\mu \pm 2\sigma)$
3σ	//	$(\mu \pm 2\sigma)$

2.2 Analysis of variance (ANOVA)

* Fisher 1930s. Agaury research England
* Analysis on linear models (*)
* Prediction vs inference (*)



i) ANOVA. Decompose total variance in a series of terms, and check how much of that variance comes from variation within samples, or between samples. (σ^2, σ_w^2, σ_B^2)

Example: We have n observations / quantities to measure, and we repeat the experiment in times (m samples)

	Sample 1	Sample 2	Sample 3
obs 1	3	5	5
obs 2	2	3	6
obs 3	1	4	7

Total average

$$\bar{x} = \frac{1}{N} \sum_{i,j}^{n,m} x_{ij} = 4 \quad (*)$$

$\bar{x}_1 = 2 \quad \bar{x}_2 = 4 \quad \bar{x}_3 = 6$

* ANOVA: analysis of variance (within and between samples)

	Sample 1	Sample 2	Sample 3
obs 1	3	5	5
obs 2	2	3	6
obs 3	1	4	7
	$\bar{x}_1 = 2$	$\bar{x}_2 = 4$	$\bar{x}_3 = 6$

Total average

$$\bar{x} = \frac{1}{N} \sum_{i,j}^{n,m} x_{ij} = 4 \quad (*)$$

i) Sum of squares total (SST) ~ variance σ^2

$$SST = \sum_{\forall i,j} (x_{ij} - \bar{x})^2 = (3-4)^2 + (2-4)^2 + (1-4)^2 + (5-4)^2 + (3-4)^2 + (4-4)^2 + (5-4)^2 + (6-4)^2 + (7-4)^2 = 30 \quad (*)$$

ii) Sum of squares within (SSW)

$$SSW = \sum_{\forall i,j} (\bar{x}_{ij} - \bar{x}_j)^2 = (3-2)^2 + (2-2)^2 + (1-2)^2 + (5-4)^2 + (3-4)^2 + (4-4)^2 + (6-6)^2 + (5-6)^2 + (7-6)^2 = 6 \quad (*)$$

iii) Sum of squares between (SSB)

$$SSB = m \sum_{\forall j} (\bar{x}_j - \bar{x}) = 3(2-4)^2 + 3(4-4)^2 + 3(6-4)^2 = 24 \quad (*)$$

* Conclusion : out of the total variance $\sigma^2 = 30$ (SST),

6 (SSW) comes from variation within the samples themselves,

24 (SSB) // between the samples.

* SST = SSW + SSB $\not\Rightarrow$ interaction term ✓ easy

We could have

SST $\neq$ SSW + SSB $\Rightarrow$ interaction (*) complicated.

(2.3) Comparing distributions. χ^2 test.

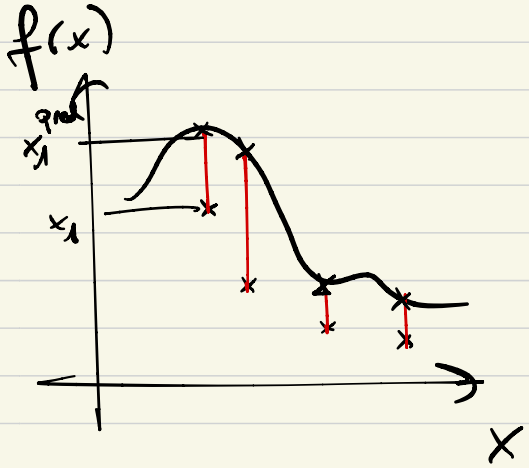
Check if a series of observations are similar among each other, or if the differences between expected values and the ones actually measured arised out of chance.

* Contingency table; comparing expectations / predicted vs measured.

Expected / predicted	Observed / measured
x_1^{pred}	x_1^{obs}
x_2^{pred}	x_2^{obs}
x_3^{pred}	x_3^{obs}

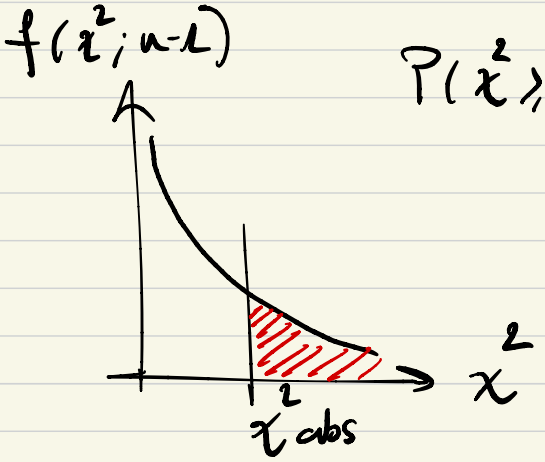
* χ^2 variable (χ^2 "statistic")

$$\chi^2 = \sum_{i=1}^{n} \frac{(x_i^{pred} - x_i^{obs})^2}{x_i^{obs}}$$



* Under the assumptions of same distributions, the χ^2 variable follows a χ^2 distribution with $n-1$ d.o.f. (*)

* Assuming H_0 true; what was probability of obtaining χ^2_{obs}?



$$P(\chi^2 \geq \chi^2_{obs}) = 1 - cdf_{\chi^2}(\chi^2_{obs}) \quad (*)$$

$(*)$ { χ^2 tables / Simulations }

r) Compare with significance level.

if p-value $> \alpha$; it was likely to obtain this obs.
We accept H_0. ✓

if p-value $< \alpha$; it was unlikely to obtain this obs.
We reject H_0. $\bar{x}_1$ and $\bar{x}_2$ different distributions.

3.1 Multiple hypothesis testing

i) P-values and significance.

When doing hypothesis testing, we could encounter errors

* type I error : incorrectly reject H_0 when it was actually true.
  (finding a false positive.)
* type II error : incorrectly accept H_0 when it was actually false.
  (finding a false negative.)

* Family-wise error rate (FWER)

"Probability of at least 1 false positive when multiple comparisons are being tested"

"Group-wise error rate". John Tukey ; 1953.

* Meaning of p-values.

We always set some significance threshold α on our p-values, to accept / reject H_0.

p-value = 0.05 ; there was a 0.05 probability of finding this result for any statistic that could have happened by chance (FP / type I error) 5 out of each 100 times.

* In large datasets / multiple hypotesis testing :

10.000 experiments ; 10.000 genes, etc.

5 % of 10.000 = 500 FP

↳ "500 genes would appear significantly different when they're not"

p-values need to be adjusted

ii) Correcting methods

* Bonferroni correction. Carlo Bonferroni (1936)

* Benjamini-Hochberg correction (1995)