HYPOTHESIS TESTING III

Sorana D. Bolboacă

OUTLINE

- ANOVA test
- Chi-square test
- Tests on proportions

Hypothesis Test: **PHANTOMS.**

Parameter
Hypotheses
Assumptions
Name the test
Test statistic
Obtain p-value
Make decision
State conclusion in context

THREE OR MORE MEANS: ANOVA

Are the means of k groups different?

- H_0: There are no differences among the m_i.
- H_1: A difference exists somewhere among the groups

Is the t-test appropriate?

No, because the t-test compare two groups and this approach will increase the size of the error to 9.75% instead of 5%

THREE OR MORE MEANS: ANOVA

Solution: apply ANOVA (analysis of variance, one-factor ANOVA or one-way ANOVA)

ANOVA Assumptions:

❶ data are independent from each other;

❷ distribution of each group in original data is normal;

❸ the variances are not significantly different by each other

THREE OR MORE MEANS: ANOVA

- Hypotheses: H_0: There are no differences among means vs. H_1: There are one or more differences somewhere among means
- Verify assumptions: ❷ normal distribution; ❸ not statistically different variances
- $\alpha = 0.05$ – df = k-1 (numerator) and df = n-k (denominator)
- F = MSM/MSE
- If $F > F_{crit} \rightarrow$ reject H_0; $F < F_{crit} \rightarrow$ failed to reject H_0

Source of variability	Sum of Squares			Mean of squares	
	Abb	Formula	df	Abb	Formula
Mean	SSM	$\sum^k n_i \cdot (m_i - m)^2$	k-1	MSM	SSM/(k-1)
Error	SSE	SST – SSM	n-k	MSE	SSE/(n-k)
Total	SST	$\sum^n (x_i - m)^2$	n-1	MST	SST/(n-1)

THREE OR MORE MEANS: ANOVA - PROBLEM

α=0.05; df: k-1=3-1=2; n-k=301-3=298; F_{crit} = 3.03; m = 66.8;
SST = 19670.3; MST = 65.57

PSA	CaP risk group	n	m	s
< 4 ng/ml	Low	89	66.1	9.1
4−10 ng/ml	Uncertain	164	66.3	7.8
>10 ng/ml	High	48	69.6	6.4

PSA=prostate-specific antigen; CaP = prostate cancer; n = sample size; m = mean; s = standard deviation

Source of variability	Sum of Squares			Mean of squares	
	Abb	Formula	df	Abb	Formula
Mean	SSM	$\sum^{k} n_i \cdot (m_i - m)^2$	k-1	MSM	SSM/(k-1)
Error	SSE	SST – SSM	n-k	MSE	SSE/(n-k)
Total	SST	$\sum^{n} (x_i - m)^2$	n-1	MST	SST/(n-1)

THREE OR MORE MEANS: ANOVA - PROBLEM

PSA	CaP risk group	n	m	s
< 4 ng/ml	Low	89	66.1	9.1
4−10 ng/ml	Uncertain	164	66.3	7.8
>10 ng/ml	High	48	69.6	6.4

PSA=prostate-specific antigen; CaP = prostate cancer; n = sample size; m = mean; s = standard deviation

$SSM = \sum^{k} n_i \cdot (m_i - m)^2$ = 89*(66.1-66.8)^2+ 164*(66.3-66.8)^2+ 48*(69.6-66.8)^2 = 460.93

MSM = SSM/2 = 460.93/2 = 230.47

SSE =SST-SSM = 19670.3 - 460.93 = 19209.37

MSE = SSE/298 = 19209.37/298 = 64.46

F = MSM/MSE = 230.47/64.46 = 3.58

Since 3.58 > 3.03 → a difference among means exists

TESTING ASSOCIATION IN CONTINGENCY TABLE

- We can perform a hypothesis test on a contingency table. The test we will use most often is the Chi-square test ($\chi 2$ test).
- χ^2 Test
  - Is proper to be applied if the sample size is large
  - The test is valid if the expected frequency of each cell is at least equal to 1 and the observed frequency is of 5
  - If the above-described conditions are not meet, the Fisher exact test is the proper test

χ^2 TEST

- Indicate if that the two variables are or are not independent BUT DO NOT quantify the power of association between them.
- Steps:
  1. Define the hypotheses
  2. Define the parameter of the test
  3. Define the significance level
  4. Define the critical interval
  5. Calculate the observed value of the parameter of the test
  6. Make a decision

χ^2 TEST: PROBLEM

- The association between *Streptococcus mutans* (as risk factor) and dental caries was studied. A sample of 620 patients was investigated. The sample contains: 150 patients with caries and *Streptococcus mutans*, 230 patients without caries and without *Streptococcus mutans* and 60 patients with caries but without *Streptococcus mutans.* The presence of *Streptococcus mutans* is asscoiated with dental caries? (df=1; α=0.05; χ^2_{critical} = 3.84).

χ^2 TEST: 1. HYPOTHESES

- H_0:
  - There is no association between *Streptococcus mutans* and dental caries.
  - The presence of *Streptococcus mutans* and dental caries are independent.
- H_1/H_a:
  - There is an association between *Streptococcus mutans* and dental caries.
  - The presence of *Streptococcus mutans* and dental caries are not independent.

χ^2 TEST: 2. PARAMETER OF THE TEST

$$\chi^2 = \sum_{i=1}^{r \cdot c} \frac{(f_i^0 - f_i^t)^2}{f_i^t}$$

Follow a distribution law with (r-1)·(c-1) degree of freedom

where

- χ^2 = the parameter of χ^2 test
- f_i^o = observed frequency
- f_i^t = expected/theoretic frequency

χ^2 TEST: 3. SIGNIFICANCE LEVEL

- Let $\alpha = 0.05$ (5%) be the significance level.

χ^2 TEST: 4. CRITICAL REGION

- Critical region: $[\chi_\alpha^2, \infty)$
- For $\alpha = 0.05$:
  - $\chi_\alpha^2 = 3.84$
  - $[3.48, \infty)$

χ^2 TEST: 5. PARAMETER OF THE TEST

observed	DC+	DC-	Total
SP +	TP = **150**	FP = **180**	330
SP -	FN = **60**	TN = **230**	290
Total	210	410	620

expected	DC+	DC-	Total
SP +	**= 330×210/620**	**= 330×410/620**	330
SP -	**= 290×210/620**	**= 290×410/620**	290
Total	210	410	620

χ^2 TEST: 5. PARAMETER OF THE TEST

observed	DC+	DC-
SP +	150	180
SP -	60	230

expected	DC+	DC-
SP +	= 112	= 218
SP -	= 98	= 192

$$\chi^2 = \frac{(150-112)^2}{112} + \frac{(180-218)^2}{218} + \frac{(60-98)^2}{98} + \frac{(230-192)^2}{192}$$

$$\chi^2 = \frac{38^2}{112} + \frac{(-38)^2}{218} + \frac{(-38)^2}{98} + \frac{(38)^2}{192}$$

$$\chi^2 = \frac{1444}{112} + \frac{1444}{218} + \frac{1444}{98} + \frac{1444}{192} = 12.89 + 6.63 + 14.73 + 7.52 = \boxed{41.77}$$

X^2 TEST: 6. MAKING DECISION

- If $\chi^2 \in [3.84, \infty)$ H_0 is rejected with a risk of error of type I (α).
- If $\chi^2 \notin [3.84, \infty)$ H_0 is accepted with a risk of error of type II (β).

- Since $41.77 \in [3.84, \infty)$ H_0 is rejected with a risk of error of 5%.
- **There is an association between *Streptococcus mutans* and dental caries.**

CONTINUITY CORRECTION (YATES'S CORRECTION)

- For small sample sizes the χ^2 test is too likely to reject the null hypothesis (it tends to spot differences where none really exist).
  - A continuity correction can be made to allow for this.
  - Two conditions have to be met:
    - All expected frequencies must be greater than 1
    - 80% of observed frequencies must be greater than 5

Yates's correction

$$\chi^2 = \sum_{i=1}^{r \cdot c} \frac{| f_i^0 - f_i^t |^2 - 0.5}{f_i^t}$$

FISHER'S EXACT TEST

- Chi-square procedures can be legitimately applied only if all values of **E** are equal to or greater than 5.
- If a 2×2 contingency table fails to meet the conditions required for the χ^2 test then Fisher's exact test can be used.
- It is based on different mathematics to the χ^2 test which are more robust when sample sizes are small.

FISHER'S EXACT TEST

- H_0: there is no association between smoking and dental caries
- If the null hypothesis is true - if any ostensible association between smoking and dental caries were the result of nothing more than mere chance coincidence -how likely is it that we might end up with a result this large or larger?

observed	DC+	DC-	Total
smoking +	TP = 2	FP = 7	9
smoking -	FN = 8	TN = 2	10
Total	10	9	19

FISHER'S EXACT TEST

- Suppose that the initial assessment was performed and the number of subjects who do and do not show characteristics (smoking and dental caries) were counted, but have not yet sorted the subjects according to the correspondences of smoking and dental caries. In this case, all they would have would be the marginal totals shown in the following table/
- Given these marginal totals, there are 10 possible ways in which the specific correspondences between smoking and dental caries.

	DC+	DC-	Total
smoking +			9
smoking -			10
Total	10	9	19

FISHER'S EXACT TEST

- The p-value is calculated directly from the formula:

$$p = \frac{(a+c)!(b+d)!(c+d)!(a+b)!}{n!a!b!c!d!}$$

- The p-value for the observed contingency table must be added to the p-value of the more extreme contingency table.

FISHER'S EXACT TEST

Obs	DC+	DC-	Total
smoking +	6	2	8
smoking -	1	6	7
Total	7	8	15

Exp	DC+	DC-	Total
smoking +	7	1	8
smoking -	0	7	7
Total	7	8	15

- The p-value must be calculated for the two contingency tables:

$$p_1 = \frac{7!8!7!8!}{15!6!2!6!} = 0.0305 \qquad p_2 = \frac{7!8!7!8!}{15!7!0!7!} = 0.0012$$

- Therefore $p = p_1 + p_2 = 0.0305 + 0.0012 = 0.0317$

- The p-value = 0.0317 < α = 0.05 $\Rightarrow$ smoking is associated with dental caries.

http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3663126/pdf/WJC-5-124.pdf

AIM: To determine whether there are gender differences in the epidemiological profile of atrial fibrillation (AF) and to characterise the clinical, biochemical, and therapeutic factors associated with AF.

Gender differences

Table 2 shows the differences between the men and the women in epidemiological, biochemical, BP and therapeutic characteristics. The women exhibited more obe-

Table 2 Epidemiological, clinical and therapeutic differences between genders (mean ± SD) *n* (%)

	Females (*n* = 542)	Males (*n* = 486)	*P* value
Mean age, yr	72.7 ± 5.8	72.8 ± 5.8	NS
Abdominal circumference, cm	96.6 ± 11.8	100.4 ± 11.0	< 0.001
Weight, kg	71.4 ± 11.5	79.5 ± 11.5	< 0.001
Mean height, cm	155.2 ± 6.7	166.7 ± 6.7	< 0.001
BMI	29.6 ± 4.5	28.6 ± 3.6	< 0.001
Obesity	224 (41.4)	160 (32.9)	0.005
Years from the onset of HT	11.0 ± 8.2	10.8 ± 8.1	NS
Diabetes mellitus	134 (24.7)	150 (30.9)	0.03
Dyslipidaemia	267 (49.3)	230 (47.3)	NS
Smokers	17 (3.1)	76 (15.6)	< 0.001
Sedentariness	352 (70.5)	274 (56.4)	< 0.001
Regular alcohol intake	5 (0.9)	33 (6.8)	< 0.001

http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3663126/pdf/WJC-5-124.pdf

AIM: To determine whether there are gender differences in the epidemiological profile of atrial fibrillation (AF) and to characterise the clinical, biochemical, and therapeutic factors associated with AF.

Table 4 Treatment differences between genders in patients with atrial fibrillation (n = 106) n (%)

	Females (n = 50)	Males (n = 56)	P value
Diuretics	34 (68.0)	30 (53.6)	NS
Beta-blockers	16 (32)	17 (30.4)	NS
Calcium antagonists	15 (30)	5 (8.9)	0.007
ACEI	12 (24)	14 (25)	NS
ARB	32 (64)	31 (55.4)	NS
Antiplatelet agents	4 (8)	16 (26.8)	0.010
VKA	29 (58)	25 (44.6)	NS
ATG or VKA	33 (66)	41 (71.4)	NS

ACEI: Angiotensin-converting enzyme inhibitors; ARB: Angiotensin receptor blocker; VKA: Vitamin K antagonist; ATG: Anti-aggregants. NS: No significant.

TESTS ON PROPORTIONS

Two-Sample Test for Binomial Proportions (Normal-Theory Test)

To test the hypothesis H_0: $p_1 = p_2$ vs. H_1: $p_1 \neq p_2$, where the proportions are obtained from two independent samples, use the following procedure:

(1) Compute the test statistic

$$z = \frac{|\hat{p}_1 - \hat{p}_2| - \left(\frac{1}{2n_1} + \frac{1}{2n_2}\right)}{\sqrt{\hat{p}\hat{q}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

where $\hat{p} = \frac{n_1\hat{p}_1 + n_2\hat{p}_2}{n_1 + n_2} = \frac{x_1 + x_2}{n_1 + n_2}, \hat{q} = 1 - \hat{p}$

and x_1, x_2 are the number of events in the first and second samples, respectively.

TESTS ON PROPORTIONS

(2) For a two-sided level α test,

if $z > z_{1-\alpha/2}$

then reject H_0;

if $z \leq z_{1-\alpha/2}$

then accept H_0.

(3) The approximate p-value for this test is given by

$$p = 2[1 - \Phi(z)]$$

(4) Use this test only when the normal approximation to the binomial distribution is valid for each of the two samples—that is, when $n_1\hat{p}\hat{q} \geq 5$ and $n_2\hat{p}\hat{q} \geq 5$.

TESTS ON PROPORTIONS – PROBLEM

The set of women with at least one birth was arbitrarily divided into two categories: (1) women whose age at first birth was ≤29 years and (2) women whose age at first birth was ≥30 years. The following results were found among women with at least one birth: 683 of 3220 (21.2%) women with breast cancer (case women) and 1498 of 10,245 (14.6%) women without breast cancer (control women) had an age at first birth ≥30. How can we assess whether this difference is significant or simply due to chance?

- p_1 = the probability that age at first birth is ≥30 in case women with at least one birth (=683/3220 = 0.212) and p_2 = the probability that age at first birth is ≥30 in control women with at least one birth (=1498/10,245 = 0.146).
- H_0: $p_1 = p_2 = p$ vs. H_1: $p_1 \neq p_2$ for some constant p.
- p = (683 + 1498)/(3220 + 10,245) = 2181/13,465 = 0.162 ⇒ q = 1 – 0.162 = 0.838
- Since $n_1pq = 3220 \cdot (0.162) \cdot (0.838) = 437 \geq 5$ and n_2pq = $10{,}245 \cdot (0.162) \cdot (.838) = 1391 \geq 5$ ⇒ z test is proper to be use

TESTS ON PROPORTIONS – PROBLEM

The set of women with at least one birth was arbitrarily divided into two categories: (1) women whose age at first birth was ≤29 years and (2) women whose age at first birth was ≥30 years. The following results were found among women with at least one birth: 683 of 3220 (21.2%) women with breast cancer (case women) and 1498 of 10,245 (14.6%) women without breast cancer (control women) had an age at first birth ≥30. How can we assess whether this difference is significant or simply due to chance?

- $$z = \frac{|0.212-0.146|-\left(\frac{1}{2\cdot 3220}+\frac{1}{2\cdot 10245}\right)}{\sqrt{0.162\cdot 0.838\cdot\left(\frac{1}{3220}+\frac{1}{10245}\right)}} = \frac{0.0657}{0.00744}, \; z = 8.8$$
- p = 1.53E-18
- Therefore, we can conclude that women with breast cancer are significantly more likely to have had their first child after age 30 than are comparable women without breast cancer.

TESTS BY EXAMPLE

Table 3. Serum and urine NGAL - paired comparisons Wilcoxon test

AKI	Wilcoxon test	NGALs1-NGALs0	NGALs2-NGALs0	NGALs2-NGALs1	NGALu1-NGALu0	NGALu2-NGALu0	NGALu2-NGALu1
Positive	Statistics	-2.429[a]	-1.955[a]	-1.122[b]	-0.653[a]	-0.296[b]	-1.009[b]
	p-value	0.015	0.051	0.262	0.514	0.767	0.313
Negative	Statistics	-3.153[a]	-2.323[a]	-0.660[b]	-0.569[b]	-0.170[a]	-1.023[a]
	p-value	0.002	0.020	0.509	0.570	0.865	0.307

NGALs0 = NGAL serum baseline; NGALs1 = NGAL serum 6h; NGALs2 = NGAL serum 12h;
NGALu0 = NGAL urinar baseline; NGALu1 = NGAL urinar 6h; NGALu2 = NGAL urinar 12h;
a = negative ranks; b = positive ranks;

Mihály O, Bolboacă SD, Rahaian R, Bodolea C, Chira C, Cristea T, Oblezniuc A, Mihály ZA, Coman I. Accuracy of Neutrophil Gelatinase-Associated Lipocalin in Detecting Acute Kidney Injury after Urogenital Robotic Assisted Laparoscopic Surgery under General Anesthesia. Applied Medical Informatics 2012;30(2):47-56.

TESTS BY EXAMPLE

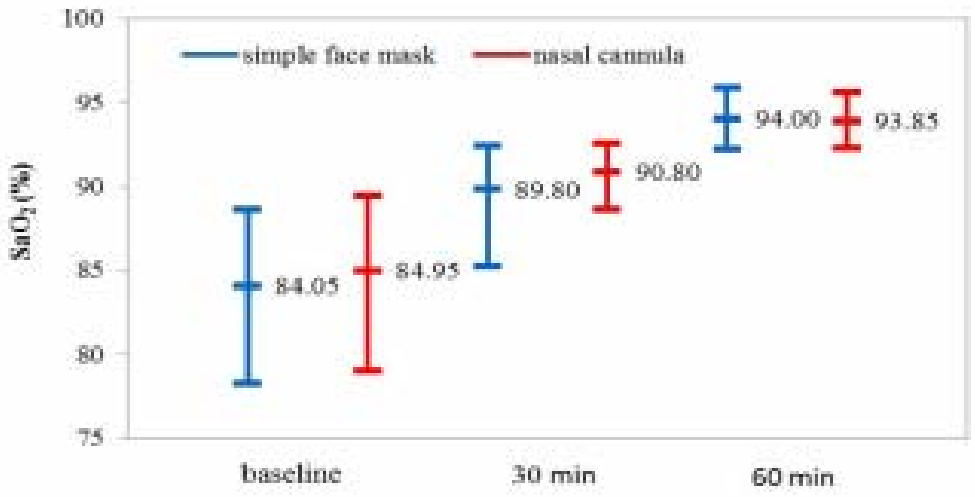

Fig. 2. SaO_2 in moment 0 (baseline), and 30 minutes and 60 minutes after oxygen therapy by simple face mask (blue) and by nasal cannula (red). The middle points represent the value of median while the extreme points are the 25th and 75th percentiles

Tab. 3. Comparison of SaO_2 and SpO_2 in the subjects who were delivered oxygen therapy by simple face mask and nasal cannula

			p-value associated to Sign test		
Group	Variable	Friedman ANOVA (p-value)	Baseline vs. 30 minutes	Baseline vs. 60 minutes	30 minutes vs. 60 minutes
Simple face mask	SaO_2 (%)	66.36 (< 0.001)	$1.36 \cdot 10^{-4}$	$1.36 \cdot 10^{-4}$	$8.64 \cdot 10^{-4}$
	SpO_2 (%)	74.05 (< 0.001)	$1.95 \cdot 10^{-4}$	$1.95 \cdot 10^{-4}$	$1.36 \cdot 10^{-4}$
Nasal cannula	SaO_2 (%)	64.39 (< 0.001)	$5.43 \cdot 10^{-4}$	$5.43 \cdot 10^{-4}$	$5.43 \cdot 10^{-4}$
	SpO_2 (%)	72.00 (< 0.001)	$3.80 \cdot 10^{-4}$	$5.43 \cdot 10^{-4}$	$1.34 \cdot 10^{-4}$

Badiu Tişa ID, Bolboacă S, Miu N, Iacob D. Efficiency of Oxygen Therapy by Simple Face Mask and Nasal Cannula for Acute Respiratory Failure in Infants and Young Children. Notulae Scientia Biologicae 2013;5(4):407-411.

TESTS BY EXAMPLE

Table 3. Sonoelastographic parameters for normal distributed variable: results of comparison between case and control group

Param	t-test for Equality of Means						
						95% CI of the Difference	
	t	df	p	MeanDiff [95%CI]	StdErrDiff	Lower	Upper
AvgRed	1.640	373.043	0.102	1.95	1.19	-0.39	4.29
AvgGreen	-2.506	257.183	$1.29 \cdot 10^{-2}$	-3.35	1.34	-5.98	-0.72
DispHue	4.04	261.76	$6.93 \cdot 10^{-5}$	3.76	0.93	1.93	5.59

t = t-value; df = degrees of freedom; MeanDiff = mean of difference;
95%CI = 95% confidence interval for mean difference; StdErrDiff = standard error of difference

Table 4. Sonoelastographic parameters for not normal distributed variable: results of comparison between case and control group

		AvgBlue	AvgIntensity	AvgHue	DispRed	DispGreen	DispBlue	DispIntensity
Most Extreme Differences	Absolute	0.429	0.230	0.694	0.253	0.348	0.106	0.176
	Positive	0.031	0.062	0.156	0.253	0.000	0.079	0.044
	Negative	-0.429	-0.230	-0.694	-0.027	-0.348	-0.106	-0.176
Kolmogorov-Smirnov Z		4.088	2.187	6.608	2.407	3.312	1.011	1.677
p		< 0.001	$1.41 \cdot 10^{-4}$	< 0.001	$1.86 \cdot 10^{-5}$	< 0.001	0.258	0.007

Botar-Jid C, Bolboacă SD, Damian L, Dudea SM, Pantilie C, Nedevschi S, Badea R. Assessment of Sonoelastography as Diagnosis Tool of Inflammatory Myopathies. Applied Medical Informatics 2010;27(4):81-89.

TESTS BY EXAMPLE

Methods: A prospective study with a six-month follow-up was conducted on hypertensive patients with LVH and mild/ moderate essential hypertension. The patients were randomly assigned to Valsartan (80 to 160 mg/day) or Nebivolol (5 to 10 mg/day) groups. The study group consisted of 108 patients, 55 in the Valsartan group and 53 in the Nebivolol group.
Results: The range of mean systolic blood pressure (SBP) varied from 152±17 (baseline) to 132±17 mmHg (follow-up) in the Valsartan group ($p<0.001$); from 146±13 to 125±14 mmHg in the Nebivolol group ($p<0.001$). The decrease in mean diastolic blood pressure (DBP) was 9.5±2.5 mmHg in the Valsartan group and 12.3±5.0 mmHg in the Ne-bivolol group. A significant reduction in QT and corrected QT (Bazett's formula) dispersion was observed in both groups, with a slightly higher reduction in the Valsartan group. Echocardiography showed a decrease in the left ven-tricle mass (LVM) indices ($p<0.05$) in both groups with a greater reduction in the Valsartan group.

Luminita Lăţea, Ştefania L. Negrea, Sorana D. Bolboacă. Effects of valsartan and nebivolol on blood pressure, QT dispersion and left ventricular hypertrophy in hypertensive patients. Dicle Medical Journal 2010;37(2):OA81-88.

STATISTICS AND MEDICAL PRACTICE

Source: http://kingeofdremes.wordpress.com/