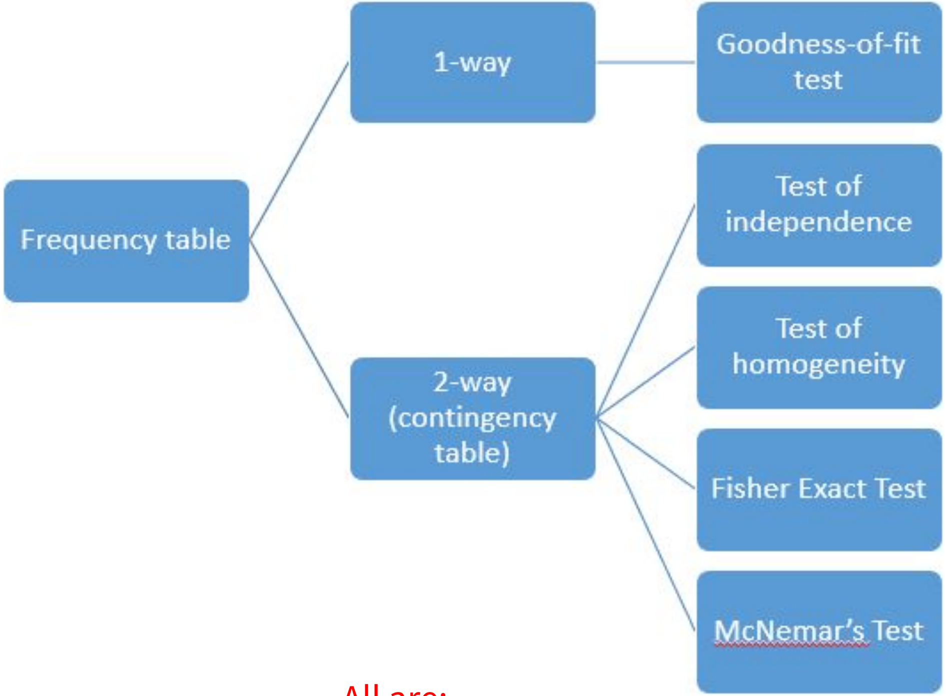
Lecture 11: Goodness-of-Fit and Contingency Tables

Applied Statistics - STAN - 5.37 & 5.38
12 & 14 January 2021
Lecturer: Erika Siregar, SST, MS

Test on Frequency Table



H0: The frequency counts agree (fit) with the claimed distribution
H1: The frequency counts do not agree (fit) with the claimed distribution

H0: The row and column variables are independent
H1: The row and column variables are dependent

H0: proporsi populasi di tiap baris sama (homogen)
H1: proporsi populasi di tiap baris tidak sama

All are:

- Right-tailed hypothesis test
- Use χ^2 for test statistics
- All except for fisher exact test require that the E >= 5

Frequency Tables

Table 11-2 Last Digits of Weights

Last Digit	Frequency
0	7
1	14
2	6
3	10
4	8
5	4
6	5
7	6
8	12
9	8

One-way frequency table

Table 11-6 Results from Experiment with Echinacea

	Treatment Group		
	Placebo	Echinacea: 20% extract	Echinacea: 60% extract
Infected	88	48	42
Not infected	15	4	10

Two-way frequency table

- 2 corresponding variables → 1 for column and another 1 for row

I. Goodness-of-Fit Test

- A test to identify whether an observed frequency distribution fits some claimed distribution.
  - Singkatnya: test untuk melihat apakah sampel yang kita pakai mengikuti distribusi tertentu berdasarkan frekuensinya.
- How to test? → hypothesis test
- Ingredients:
  - Data consists of frequency counts for several different categories → example
  - The hypothesis
    H0: The frequency counts **agree/fit** with the claimed distribution
    H1: The frequency counts **do not agree/fit** with the claimed distribution
  - Test statistics: χ^2

$$\chi^2 = \sum \frac{(O - E)^2}{E}$$

O = observed frequency → **observation**
E = expected frequency → **theoritically**
k = number of different categories
df = k - 1

  - **E = n/k** → if it's assumed that the expected frequency are **all equal**.
  - **E = np** → if it's assumed that the expected frequency are **not all equal**.
  - For each category, **E >= 5**
  - Reject H0 if **p-value ≤ α**

Table 11-2 Last Digits of Weights

Last Digit	Frequency
0	7
1	14
2	6
3	10
4	8
5	4
6	5
7	6
8	12
9	8

Example for E

a. Equally likely:
Sebuat dadu dilempar 45 kali dengan hasil seperti yang tampak pada tabel. Tentukan expected frequency (E)-nya.

Outcome	1	2	3	4	5	6
Observed freq. (O)	13	6	12	9	3	2
Expected freq (E)	?	?	?	?	?	?

→ **categories**

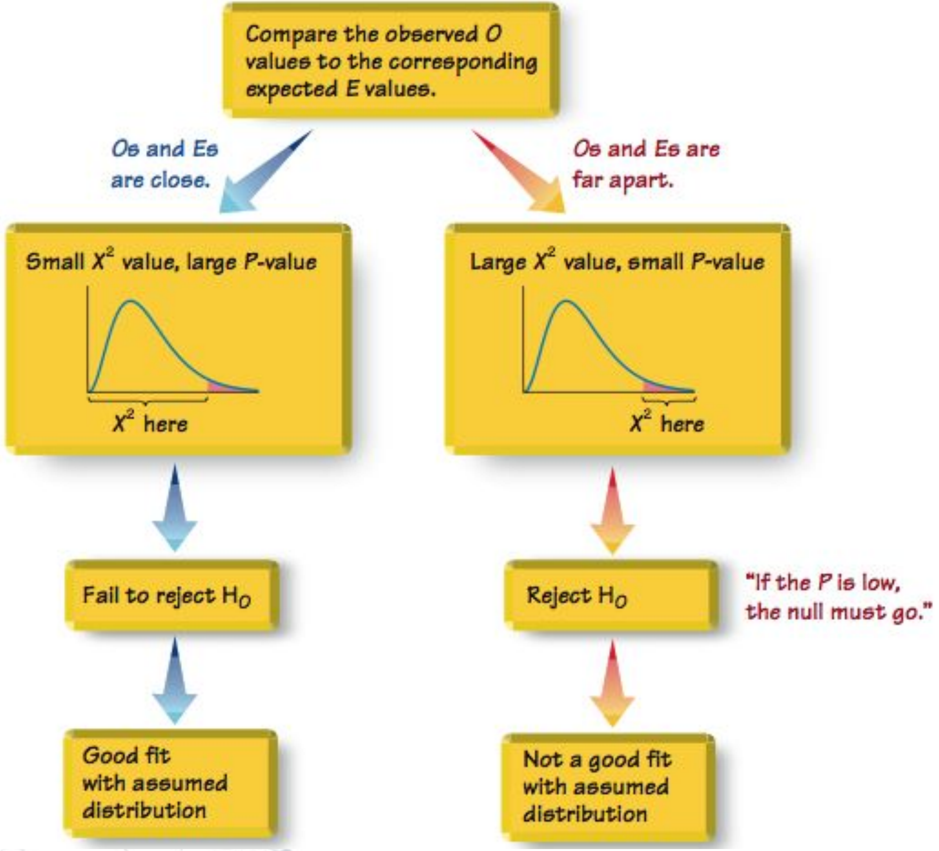
E = n/k

b. Not equally likely:
Let's say dadunya dibuat tidak seimbang, sehingga angka 1 memiliki peluang muncul 50%, sedangkan angka lainnya berpeluang muncul 10%.

Outcome	1	2	3	4	5	6
Probability	0.5	0.1	0.1	0.1	0.1	0.1
Observed freq. (O)	13	6	12	9	3	2
Expected freq (E)	?	?	?	?	?	?

E = np

Relationships Among the 2 Test Statistic, P-Value, and Goodness-of-Fit



- The test statistics χ^2 is the difference between O and E.
- **Goodness-of-fit hypothesis** tests are always **right-tailed** → critical values will always be on the right side of the curve.

Example for Goodness-of-Fit Test

last_digit	obs_freq	exp_freq
0	46	10
1	1	10
2	2	10
3	3	10
4	3	10
5	30	10
6	4	10
7	0	10
8	8	10
9	3	10

Dikumpulkan data berat badan 100 orang secara random. Researchers ingin menguji apakah data ini dikumpulkan dengan benar-benar mengukur berat 100 orang tersebut, atau hanya bertanya ke ybs saja, dimana orang cenderung membulatkan beratnya ke digit berakhiran 0 atau 5. So, with $\alpha = 5\%$, we'd like to test the claim that the digits do not occur with the same frequency. → H0: all digits have the same probability of being occured → **uniform distribution**

H0: p0 = p1 = p2 = p3 = p4 = p5 = p6 = p7 = p8 = p9
(the last digits occur with same relative frequency)

H1: At least one of the p is different from others.
(the last digits do not occur with same relative frequency)

The Manual Way

last_digit	obs_freq	exp_freq	oe	oe2	oe2e
0	46	10	36	1296	129.6
1	1	10	-9	81	8.1
2	2	10	-8	64	6.4
3	3	10	-7	49	4.9
4	3	10	-7	49	4.9
5	30	10	20	400	40
6	4	10	-6	36	3.6
7	0	10	-10	100	10
8	8	10	-2	4	0.4
9	3	10	-7	49	4.9
Jumlah	100	100	0	2128	**212.8**

Test statistics: $\chi^2 = 212.8 \rightarrow df = 9$

P-value = ...?

Critical value = ...?

Decision= ...?

Interpretation:

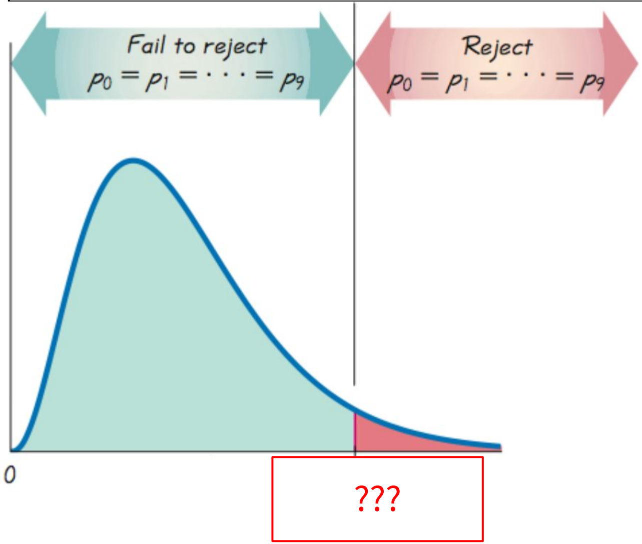
Try it using R

Using R: *chisq.test(obs_data)*

```
> weight_gof <- chisq.test(weight$obs_freq)
> weight_gof

        Chi-squared test for given probabilities

data:  weight$obs_freq
X-squared = 212.8, df = 9, p-value < 2.2e-16
```



Contingency Tables

- 2-way frequency table → tabel frekuensi 2 arah.
  - have **at least 2 rows** and **at least 2 columns**
  - frequencies correspond to two variables
  - **1st** variable → categorize **rows**
  - **2nd** variable → categorize **columns**
  - For every cell in the contingency table, the expected frequency E is at least 5

Table 11-6 Results from Experiment with Echinacea

	Treatment Group		
	Placebo	Echinacea: 20% extract	Echinacea: 60% extract
Infected	88	48	42
Not infected	15	4	10

- Hypothesis
  H0: The row and column variables are independent
  H1: The row and column variables are dependent

Contingency Tables (2)

- **Hypothesis:**
  H0: The row and column variables are independent

  H1: The row and column variables are dependent

- Test Statistics

$$\boxed{\chi^2 = \sum \frac{(O - E)^2}{E}}$$

df = (r-1)(c-1)

$$\boxed{E = \frac{(\text{row total})\,(\text{column total})}{(\text{grand total})}}$$

Derived from the ***n*p*** formula below

$$E = \underbrace{\cancel{\text{grand total}} \cdot \frac{\text{row total}}{\cancel{\text{grand total}}}}_{n} \cdot \underbrace{\frac{\text{column total}}{\text{grand total}}}_{p}$$

$n \cdot p$

(probability of a cell)

Table 11-6 Results from Experiment with Echinacea

	Treatment Group		
	Placebo	Echinacea: 20% extract	Echinacea: 60% extract
Infected	88	48	42
Not infected	15	4	10

- Tests of Independence are always **right-tailed**.

Example

$$E = \frac{(\text{row total})\ (\text{column total})}{(\text{grand total})}$$

Table 11-6 Results from Experiment with Echinacea

	Treatment Group			
	Placebo	Echinacea: 20% extract	Echinacea: 60% extract	
Infected	88	48	42	178
Not infected	15	4	10	29
	103	52	52	**207**

In a test of the effectiveness of echinacea to fight rhinovirus, some test subjects were treated with echinacea extracted with **20% ethanol**, some were treated with echinacea extracted with **60% ethanol**, and others were given a **placebo**. All of the test subjects were then exposed to **rhinovirus**. Use a 0.05 significance level to test the claim that **getting an infection (cold) is independent of the treatment group**. What does the result indicated about the effectiveness of echinacea as a treatment for colds?

- For the **1st cell (88)** → **E = 178*103/207 = 88.57**
- Interpretation:
  if we assume that getting an infection is independent of the treatment, then we expect to find that 88.57 of the subjects would be given a placebo and would get an infection.
- O - E → 88 - 88.57 → key component of the test statistics.

Example

Table 11-6 Results from Experiment with Echinacea

	Treatment Group			
	Placebo	Echinacea: 20% extract	Echinacea: 60% extract	
Infected	88 88.57	48	42	178
Not infected	15	4	10	29
	103	52	52	**207**

$$E = \frac{(\text{row total})(\text{column total})}{(\text{grand total})}$$

In a test of the effectiveness of echinacea to fight rhinovirus, some test subjects were treated with echinacea extracted with **20% ethanol**, some were treated with echinacea extracted with **60% ethanol**, and others were given a **placebo**. All of the test subjects were then exposed to **rhinovirus**. Use a 0.05 significance level to test the claim that **getting an infection (cold) is independent of the treatment group**. What does the result indicated about the effectiveness of echinacea as a treatment for colds?

Answer:

H0: Getting an infection is independent of the treatment

H1: Getting an infection and the treatment are dependent

Test statistics

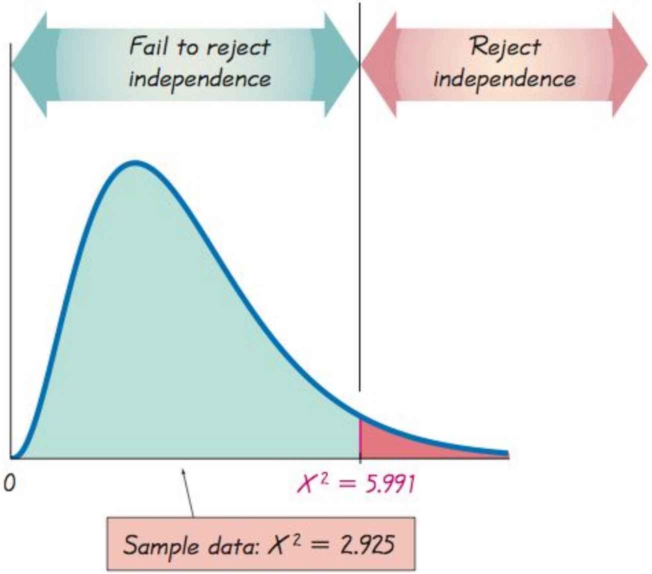
$$\chi^2 = \sum \frac{(O - E)^2}{E}$$

$$= \frac{(88 - 88.570)^2}{88.570} + \ldots + \frac{(10 - 7.285)^2}{7.285}$$

$$= 2.925$$

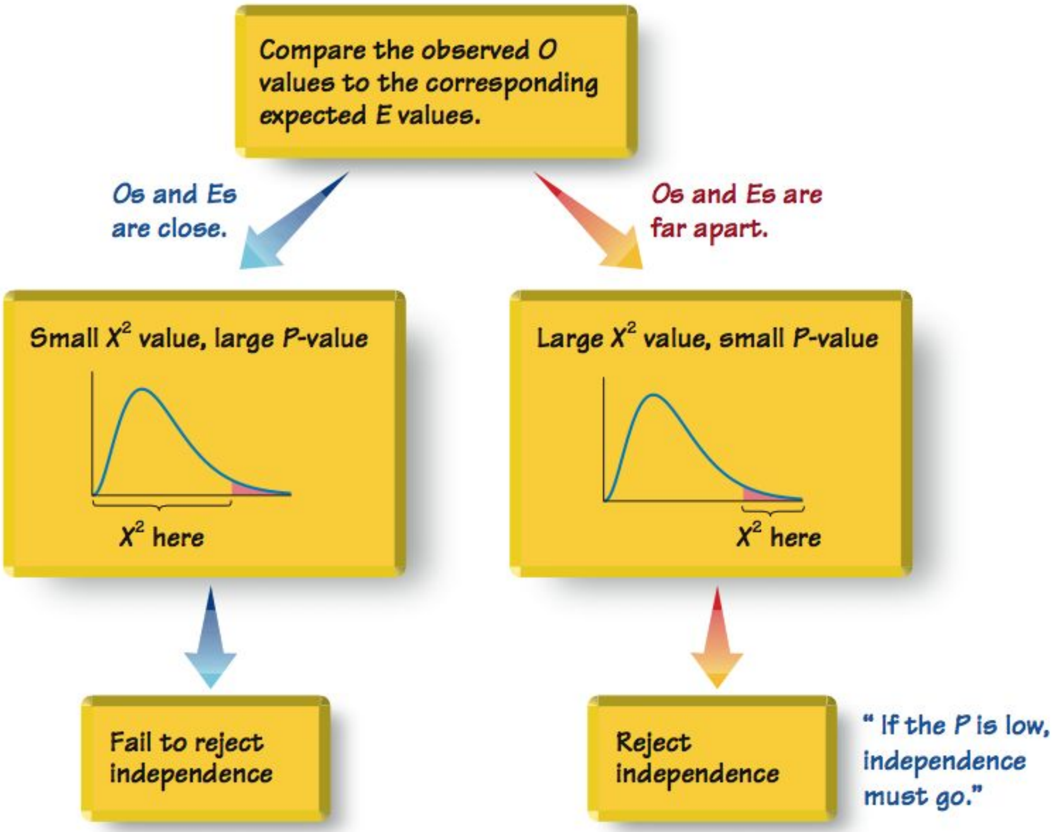
df = (r-1)(c-1) = (2-1)(3-1) = 2

For the **1st cell (88)** → **E = 178*103/207 = 88.57**
Interpretation:
if we assume that getting an infection is independent of the treatment, then we **expect to find that 88.57** of the subjects would be given a placebo and would get an infection.
O - E → 88 - 88.57 → key component of the test statistics.



- Test statistics = 2.925
- P-value =
  > pchisq(2.925, 2, lower.tail = FALSE)
  [1] **0.2316564**
- Critical value = with α=5% & df = 2
  > qchisq(0.05, 2, lower.tail = FALSE)
  [1] **5.991465**
- P-value > α → fail to reject H0.
- Interpretation:
  Getting an infection is **independent** of the treatment group. This suggests that **echinacea is not an effective treatment for colds.**

Relationships Among Key Components in Test of Independence



Test of Homogeneity

- Claim: **different populations** have the **same proportions** of some characteristics
- **Hypothesis:**
  H0: proporsi populasi di tiap baris sama (homogen)
  H1: proporsi populasi di tiap baris tidak sama
- E and test statistics == test of independence.

$$\chi^2 = \sum \frac{(O - E)^2}{E}$$

df = (r-1)(c-1)

$$E = \frac{(\text{row total})(\text{column total})}{(\text{grand total})}$$

- Independency Test vs Homogeneity Test:
  - **Sample size**: pada **homogeneity test**, jumlah sampel untuk tiap proporsi (kolom/baris) sudah **ditentukan dari awal**, sementara pada **independency test** kita menarik 1 set sampel besar, kemudian **secara random dihitung** berapa yang masuk ke masing-masing kolom/baris.
  - **Tujuan:**
    - **Homogeneity** test: **apakah proporsi** populasi di tiap baris **sama**?
    - **Independence** test: **apakah ada keterkaitan** antara column variable dengan row variable?

Example

H0: The proportions of agree/disagree responses are the same for the subjects interviewed by men and the subjects interviewed by women.
H1: The proportions are different.

A group of surveyed men were asked if they agreed with this statement: "*Abortion is a private matter that should be left to the woman to decide without government intervention.*" Assume that the survey was designed so that **male interviewers** were instructed to obtain **800 responses** from male subjects, and **female interviewers** were instructed to obtain **400 responses** from male subjects. Using a **0.05** significance level, test the claim that the **proportions of agree/disagree** responses are the **same** for the subjects **interviewed by men** and the subjects **interviewed by women**.

Table 11-6 Gender and Survey Responses

	Gender of Interviewer	
	Man	Woman
Men who agree	560	308
Men who disagree	240	92

```
> abortion_chisq

	Pearson's Chi-squared test with Yates' continuity correction

data:  abortion
X-squared = 6.1842, df = 1, p-value = 0.01289
```

P-value < α → reject H0
Meaning: **no homogeneity**

Fisher Exact Test

- for a 2 x 2 contingency table.
- Having 1 or more expected frequencies < 5
- calculations are quite complex, it's a good idea to use technologies (R, excel, minitab, spss, etc).
- With R:
  fisher.exact(data)

McNemar's Test for Matched Pair

- testing the null hypothesis that the **frequencies from the discordant (different) categories occur in the same proportion**.
- Data are place in a 2 x 2 table where each observation is classified in two ways.
- The test only compares categories that are **different** (**discordant** pairs).
- **Table 11-9** 2 × 2 Table with Frequency Counts from Matched Pairs

		Treatment X	
		Cured	Not Cured
Treatment Y	Cured	*a*	$\boxed{b}$
	Not cured	$\boxed{c}$	*d*

use only the frequencies from the pairs of categories that are different

a, b, c, and d represent the frequency counts

- H0: The proportions of the frequencies b and c are the same.
  H1: The proportions of the frequencies b and c are different.
- Test statistics: $$\chi^2 = \frac{(|b - c| - 1)^2}{b + c}$$

Example

A randomized controlled trial was designed to **test the effectiveness of hip protectors** in preventing hip fractures in the elderly. Nursing home residents each **wore protection on one hip, but not the other (matched pair)**. Results are summarized in table. Using a **0.05** significance level, apply McNemar's test to test the null hypothesis that the following two proportions are the same:

		No Hip Protector Worn	
		No Hip Fracture	Hip Fracture
Hip Protector Worn	No Hip Fracture	309	10
	Hip Fracture	15	2

- The proportion of subjects with **no hip fracture on the protected hip** and **a hip fracture on the unprotected hip**.
- The proportion of subjects with a **hip fracture on the protected hip** and **no hip fracture on the unprotected hip**.

Based on the results, **do the hip protectors appear to be effective in preventing hip fractures**?

$$\chi^2 = \frac{(|b - c| - 1)^2}{b + c} = \frac{(|10 - 15| - 1)^2}{10 + 15} = 0.640$$

```
> mcnemar.test(hip)

	McNemar's Chi-squared test with continuity correction

data:  hip
McNemar's chi-squared = 0.64, df = 1, p-value = 0.4237
```

It appears that the **proportion of hip fractures with the protectors worn** is **<u>not significantly different</u>** from the proportion of **hip fractures without the protectors worn**. The hip protectors **do not appear to be effective** in preventing hip fractures.

Exercise (Note: all tests use α = 5%)

1. Lakukan goodness-of-fit test pada data internet (internet.csv), definisikan hipotesisnya, hitung test statistiknya, buat keputusan dan interpretasi.
   a. Dengan cara manual
   b. Dengan R
2. Berdasarkan tabel berikut, ujilah independensi antara “the type of treatment for stress facture” vs ‘its success’.

	success	failure
surgery	54	12
Weight-bearing cast	41	51
non-weight-bearing cast for 6 weeks	70	3
non-weight-bearing cast for < 6 weeks	17	5

Exercise

3. Berdasarkan tabel berikut, lakukanlah uji homogenitas dengan $\alpha = 5\%$

UU Baru	Partai politik			Total
	Golkar	PDI-P	Demokrat	
Setuju	82	70	62	**214**
Menentang	93	62	67	**222**
Tanpa pendapat	25	18	21	**64**
Total	**200**	**150**	**150**	**500**

Exercise

4. Lakukan fisher's exact test berdasarkan data berikut untuk menguji hipotesis tentang apakah ada hubungan antara melihat orang menguap dengan membuat kita jadi ikutan menguap (yawning)

		Subject exposed to yawning?	
		yes	no
Did subject yawn?	Yes	10	4
	No	24	12

5. Lakukan mcnemar test pada data berikut. Sertakan juga hasil penghitungan test statisticsnya dengan cara manual.

1st survey		2nd survey	
		approve	disapprove
1st survey	approve	794	150
	disapprove	86	570

Thanks!

Any questions?

You can find me at:

@erikaris