STAC51 TUT02

Midterm Review

Week 7: Feb 25, 2021

Question 2

2 pts

True or False?

For testing independence with random samples, Pearson's X^2 statistic and the likelihood-ratio G^2 statistic both have chi-squared distributions for any sample size, as long as the sample was randomly selected.

Large N.

- True
- False

X \ Y	NO	YES	
Treatment			n_1
control			n_2
	n_1	n_2	

$\pi_1 = p(\text{YES} \mid \text{Treatment})$ ~~control~~

$\frac{\pi_1}{1-\pi_1} \Big/ \frac{\pi_2}{1-\pi_2}$

$p(\text{Treatment} \mid \text{YES})$

$\delta_1 =$

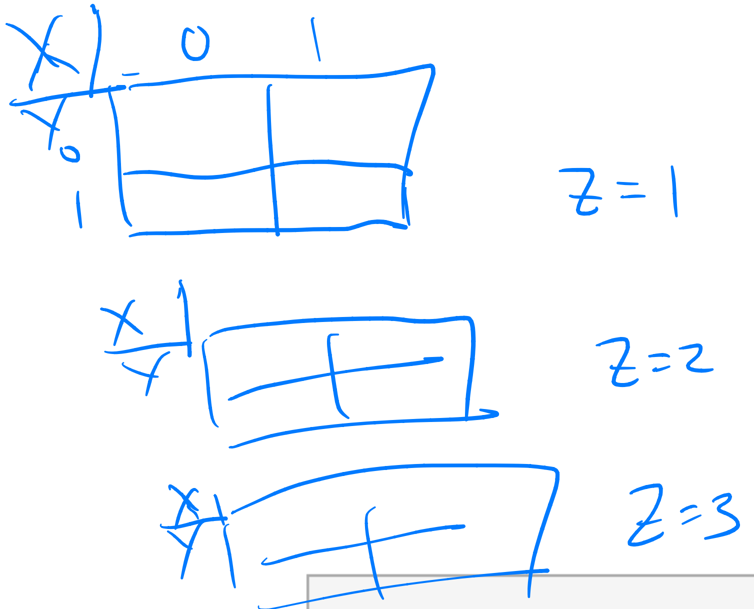
Question 4

2 pts

True or False?

The difference of proportions, relative risk, and odds ratio are valid measures for summarizing 2×2 tables for either prospective or retrospective (e.g., case-control) studies.

- True
- False



$\Leftrightarrow$ $OR(X, Y \mid Z) \neq OR(X, Y)$.

θ

Question 6

2 pts

Suppose X and Y are two 2-level categorical random variables and Z is a 3-level categorical random variable. If the conditional odds-ratios between X and Y given Z are the same (and denoted by θ) for all levels of Z, then we can conclude that

Homogeneous Association

- the odds-ratio between X and Y is equal to θ.
- None of the above
- X and Y are independent. $\Rightarrow$ $OR(X, Y) = 1$
- X and Y are conditionally independent given Z. $\theta = OR(X, Y \mid Z) = 1$

How do we test for conditionally Independent?

CMH Test.

Question 7

4 pts

For adults who sailed on the Titanic on its fateful voyage, the odds ratio between gender (~~female, male) and sur~~vival (yes, no) was 10. The odds of survival for females was 2.5. Find the probability of males who survived.

0.2.

$$\Rightarrow \frac{\text{odds(female)}}{\text{odds (male)}} = 10$$

$$\text{odds(male)} = \frac{\text{odds(female)}}{10} = 0.25$$

$$\boxed{\frac{\pi_m}{1-\pi_m} = 0.25} \Rightarrow \pi_m = 0.2$$

$$\Lambda = 2[\ell(\hat{\mu}) - \ell(\mu_0)] \sim \chi^2_{(1)}$$

$$L(\mu) = C \prod_{i=1}^{n} \mu^{x_i} \exp(-\mu) \propto \mu^{\sum x_i} \exp(-n\mu)$$

$$\boxed{\ell(\mu)} = \sum x_i \log(\mu) - n\mu \qquad S(\mu) = \frac{\sum x_i}{\mu} - n = 0$$

$$\text{MLE}: \hat{\mu} = \boxed{\frac{\sum x_i}{n} = \bar{x}} = \frac{20}{5} = 4$$

The number of accidents at a particular crossing has a Poisson distribution with parameter, μ unknown. The accidents recorded in the last 5 months are as follows: **2, 8, 1, 4, 5**. Assume that the accidents on different months are independent. $n=5$

We would like to use the likelihood ratio test to test the hypothesis:

$H_0 : \mu = 2.5 \quad vs \quad H_1 : \mu \neq 2.5.$

Compute the likelihood ratio test statistic. Round the number to the second decimal place, for example, 3.567 is 3.57.

$$2\left[\sum(x_i) \log\left(\frac{\hat{\mu}}{\mu_0}\right) - n[\hat{\mu} - \mu_0]\right]$$

$$2 \times \Big(\underset{20}{\overset{\|}{}} \times \log\left(\frac{4}{2.5}\right) - 5[4 - 2.5] \Big) = \boxed{3.8}$$

$$L(\theta) = g(x)\, h(\theta, x)$$

$$\frac{L(\theta_1)}{L(\theta_2)} = \frac{\cancel{g(x)}\, h(\theta_1, x)}{\cancel{g(x)}\, h(\theta_2, x)}$$

$\chi^2(1)$

Question 10

4 pts

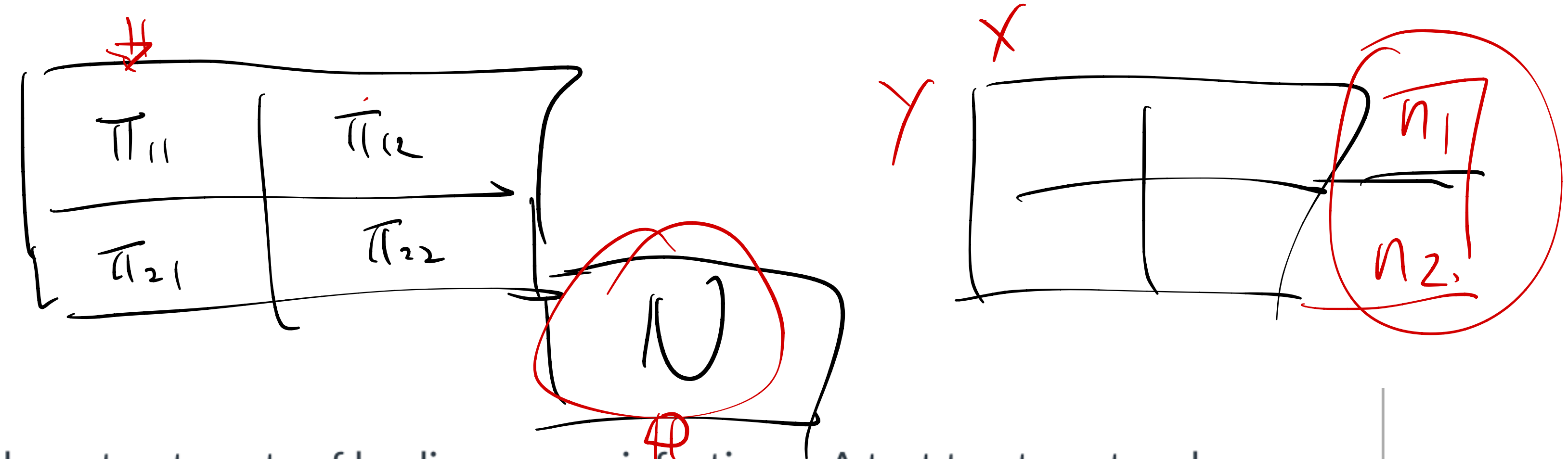
The number of accidents at a particular crossing has a Poisson distribution with parameter, μ unknown. The accidents recorded in the last 5 months are as follows: **2, 8, 1, 4, 5.** Assume that the accidents on different months are independent.

LRT = 3.8.

We would like to use the likelihood ratio test to test the hypothesis:
$H_0 : \mu = 2.5 \quad vs \quad H_1 : \mu \neq 2.5$.

Based on the likelihood ratio test statistic that you computed before, what are your p-value (use the rounded test statistic value in the previous question) and conclusion? Use $\alpha = 0.05$.

- P-value is extremely small, so we have a strong evidence to reject the null hypothesis.
- P-value is about 0.949, and we don't have enough evidence to reject the null hypothesis.
- ✓ P-value is about 0.0513, and we don't have enough evidence to reject the null hypothesis.
- P-value is about 0.1025, and we don't have enough evidence to reject the null hypothesis.
- P-value is less than 0.05, so we have enough evidence to reject the null hypothesis.



Consider the following data from a study on treatments of healing severe infections. A test treatment and control are compared to determine whether the rates of favorable responses are the same.

Treatment	Favorable	Unfavorable	Total
Test	10	2	12
Control	2	4	6
Total	12	6	18

$X_1 \sim Bin(n_1, \pi_1)$

$X_2 \sim Bin(n_2, \pi_2)$

$P(X_1 = x_1 \mid X_1 + X_2 = z)$

$\hookrightarrow$ Hypergeometric

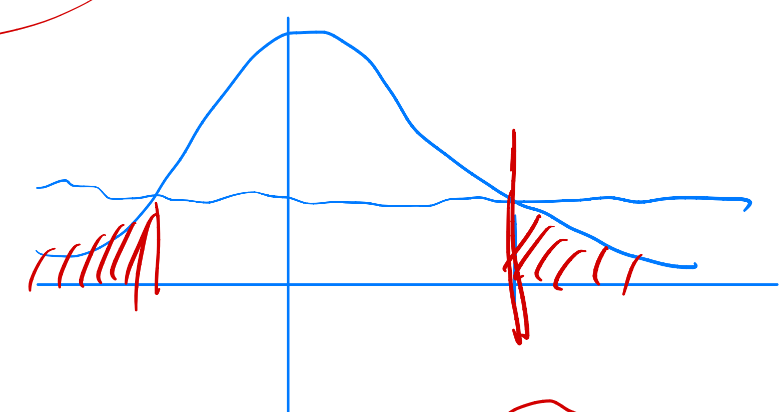
If we fix the margins (12, 6, 12, 6), what will be the distribution of cell counts?

- Hypergeometric distribution
- Binomial distribution
- Poisson distribution
- Multinomial distribution

Consider the following data from a study on treatments of healing severe infections. A test treatment and control are compared to determine whether the rates of favorable responses are the same. Compute the p-value for the two-sided test based on Fisher's exact test.

$\Theta = 1$

Treatment	Favorable	Unfavorable	Total
Test	10	2	12
Control	2	4	6
Total	12	6	18



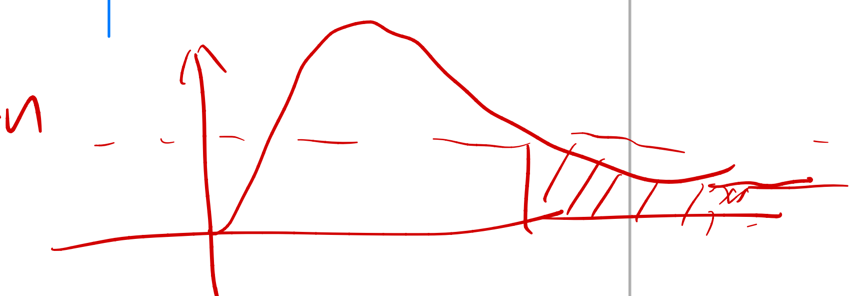
Here is the R output that you can use to compute it.

```
> round(dhyper(6:12, 12, 6, 12), 5)
[1] 0.04977 0.25598 0.39997 0.23702 0.05333 0.00388 0.00005
```

6 7 8 9 10 11 12

No greater than

$\leq$



- 0.29428
- 0.34405
- 0.05726
- ✓ 0.10703
- 0.05333

Consider the following data from a study on treatments of healing severe infections. A test treatment and control are compared to determine whether the rates of favorable responses are the same.

Treatment	Favorable	Unfavorable	Total
Test	10	2	12
Control	2	4	6
Total	12	6	18

Conduct a Pearson's chi-square test of independence **without continuity correction.** Use R function to perform the analysis and provide the test statistic value with the full decimal places.

4.5

Consider the following data from a study on treatments of healing severe infections. A test treatment and control are compared to determine whether the rates of favorable responses are the same.

Treatment	Favorable	Unfavorable	Total
Test	10	2	12
Control	2	4	6
Total	12	6	18

Conduct a G^2 likelihood ratio test of independence. Use R function to perform the analysis and provide the test statistic value **with a second decimal point.** For example, 2.3256 will be 2.33.

4.46

Question 16

2 pts

Consider the following data from a study on treatments of healing severe infections. A test treatment and control are compared to determine whether the rates of favorable responses are the same.

Treatment	Favorable	Unfavorable	Total
Test	10	2	12
Control	2	4	6
Total	12	6	18

Which test among three that we conducted previously and the conclusion would you report?

- Fisher's exact test, there is a significant association with α = 0.05.
- The likelihood ratio test G^2 , there is a significant association with α =0.05.
- Fisher's exact test, there is no significant association with α = 0.05.
- Pearson's chi-squre test, there is a significant association with α =0.05.

This is the only **theoretical question** that you can write down the solution at the paper, and scan it, make it as "**pdf file**" and upload the file.

Suppose $Y_1, Y_2, \ldots, Y_n$ is a random sample from the Poisson distribution with a mean parameter μ so that its probability mass function is $\frac{e^{-\mu}\mu^y}{y!}$. Answer the following questions:

Let's assume that we can find the MLE of μ in the previous question. I would like to find the variance of log of $\hat{\mu}$ for the testing hypothesis purpose.

1) Find the Var(log $\hat{\mu}$)) using the delta method. (6 pt)

2) Suppose we have a random sample of 10, 8, 12, 10 ~ Poisson(μ). Construct a 95% large sample CI for log(μ). (4 pt)

1) $\boxed{\operatorname{Var}(\log(\hat{\mu}))}$ $\quad g(\hat{\mu}) = \log(\hat{\mu})$

$g'(\mu) = \frac{1}{\mu}$

Delta-method : g' exist, $\neq 0$

$$\operatorname{var}[g(\hat{\theta})] = [g'(\theta)]^2 \operatorname{Var}(\hat{\theta})$$

θ $\boxed{\hat{\mu} = \bar{X}}$ $\quad \operatorname{Var}(\hat{\mu}) = \operatorname{Var}(\bar{X}) = \frac{\operatorname{Var}(X)}{n} = \frac{\mu}{n}$

$$\rightarrow [g'(\mu)]^2 \operatorname{Var}(\hat{\mu}) = \left[\frac{1}{\mu}\right]^2 \frac{\mu}{n} = \frac{1}{n\mu}$$

$$\widehat{\operatorname{Var}}(\log(\hat{\mu})) = \frac{1}{n\hat{\mu}} = \frac{1}{\sum X_i}.$$

2) $\boxed{95\%}$ C.I. $\log(\hat{\mu})$

$$\log(\hat{\mu}) \pm 1.96 \times \sqrt{\frac{1}{n\hat{\mu}}}$$

$\hat{\mu} = \bar{X}$

$$= \log(\bar{X}) = \log(10)$$

$$\boxed{\log(10) \pm 1.96 \sqrt{\frac{1}{40}}}$$

$$\Rightarrow \boxed{\operatorname{var}(\hat{\theta}) = [j(\hat{\theta})]^{-1} = -E\left(\frac{\partial^2 \ell(\theta)}{\partial \theta^2}\right)}$$

$$\widehat{\operatorname{var}}(\hat{\theta}) = [j(\theta)]^{-1}\Big|_{\theta=\hat{\theta}}.$$

2. For the 23 space shuttle flights that occurred before Challenger mission disaster in 1986, the following table shows the temperature (°F) at the time of the flight and whether at least one primary O-ring suffered thermal distress.

Ft	Temp	TD	Ft	Temp	TD	Ft	Temp	TD
1	66	0	9	57	1	17	70	0
2	70	1	10	63	1	18	81	0
3	69	0	11	70	1	19	76	0
4	68	0	12	78	0	20	79	0
5	67	0	13	67	0	21	75	1
6	72	0	14	53	1	22	76	0
7	73	0	15	67	0	23	58	1
8	70	0	16	75	0			

(a) Use logistic regression to model the effect of temperature on the probability of the thermal distress. Interpret the model fit.

(b) Calculate the predicted probability of thermal distress at 31°, the temperature at the time of the Challenger flight. At what temperature does the predicted probability equal 0.5? At that temperature, give a linear approximation for the change in the predicted probability per degree increase in temperature.

(c) Interpret the effect of temperature on the odds of thermal distress. Test the hypothesis that temperature has no effect, using (i) the Wald test, (ii) the likelihood-ratio test.