Math 2311

Bekki George – bekki@math.uh.edu

Office Hours: MW 11am to 12:45pm in 639 PGH

Online Thursdays 4-5:30pm

And by appointment

Class webpage: http://www.math.uh.edu/~bekki/Math2311.html

8.6 – Inference for Two-way Tables

not on Test3 – is on Q14 + final

We can also use the Chi-Square method to make inferences for data in two-way tables.

The formula to find the expected count in a two-way table is:

$$\text{expected count} = \frac{\text{row total} \times \text{column total}}{n}$$

Where n is the grand total of all values.

When conducting a Chi-square test of independence in a two-way table, the null and alternate hypothesis will be:

H_0: There is no association between the row and column variables.

H_a: There is an association between the two variables.

The test statistic is:

$$\chi^2 = \sum_{allcells} \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$$

The p-value: $P(\chi_k^2 > \chi^2)$, where χ_k^2 represents a Chi-square distribution with $df = (r-1)(c-1)$ degrees of freedom.

The assumptions necessary for the test to be valid are:

1. The observations constitutes a simple random sample from the population of interest, and
2. The expected counts are at least 5 for each cell of the table.

By itself, the chi-square test determines only whether the data provide evidence of a relationship between the two variables. If the result is significant, one can go on to identify the source of that relationship by finding the cells of the table that contribute most to the χ^2 value (i.e. those cells with the biggest discrepancy between the observed and expected counts) and by noting whether the observed count falls above or below the observed count in those cells.

Example:

3. Use the data below to determine if there is sufficient evidence to conclude that an association exists between car color and the likelihood of being in an accident.

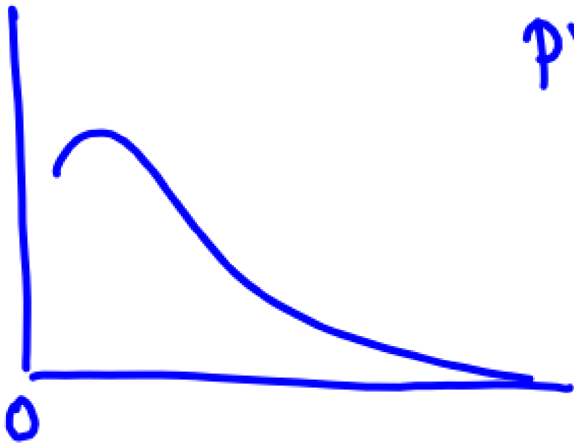
	Red	Blue	White	
Car has been in accident	28	33	36	97
Car has not been in accident	23	22	30	75
	51	55	66	$\boxed{172}$

$r = 2$
$c = 3$

$r - 1 = 1$
$c - 1 = 2$
$df = 1 \cdot 2 = 2$

Expected counts	Red	Blue	White
accd.	$\frac{97(51)}{172}$	$\frac{97(55)}{172}$	$\frac{97(66)}{172}$
no accd.	$\frac{75(51)}{172}$	$\frac{75(55)}{172}$	$\frac{75(66)}{172}$

$$\chi^2 = \frac{(28 - 28.76)^2}{28.76} + \frac{(33 - 31.02)^2}{31.02} + \frac{(36 - 37.22)^2}{37.22} + \dots = .43$$



pvalue : $p(\chi^2 > .43) = .81 > .05$

fail to reject H_0

$\chi^2 cdf(.43, 99999, 2)$

1-pchisq(.43, 2)

H_0: no association

H_a: is association

Mixed up problems! **For each of the following, (a) identify the type of test to be used, and (b) state the hypothesis.**

1. The Blue Diamond Company advertises that their nut mix contains (by weight) 40% cashews, 15% Brazil nuts, 20% almonds and only 25% peanuts. The truth-in-advertising investigators took a random sample (of size 20 lbs) of the nut mix and found the distribution to be as follows:

Cashews	*Brazil Nuts*	*Almonds*	*Peanuts*
6 lb	3 lb	5 lb	6 lb

At the 0.01 level of significance, is the claim made by Blue Diamond true?

χ^2 goodness of fit

H_0: no difference

H_a: is a difference

Ch7 - Confidence Intervals

Statistic ± Margin of error

$\bar{x}$ or $\hat{p}$

in middle of interval

z^* or t^* and st. deviation (st. error)

$\frac{1+CL}{2}$

$[8, 26]$

$\bar{x} = 17$

$ME = 9$

Finding n:

ME ±.03

95% conf.

$s = 1.7$

ME > formula for ME

$$.03 > 1.96 \cdot \frac{1.7}{\sqrt{n}}$$

$$\sqrt{n}^2 > \left[\frac{1.96(1.7)}{.03}\right]^2$$

$n = 12336$

larger conf. level → wider interval

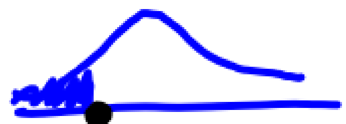
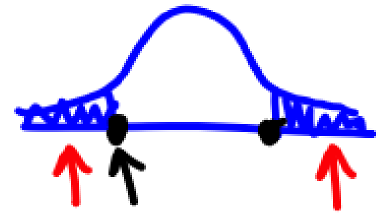
larger n → narrower interval

larges st. error → wider int.

Hyp. Testing

H_0: <see Mondays notes>

H_a: change = to

$<$	$>$	$\neq$
		
α	α	$\alpha/2$ $\alpha/2$
invnorm invT .... (α)	inv_ $(1-\alpha)$	inv_ $(\alpha/2)$

test statistic = $\square$

pvalue	$p(z \text{ or } t < \square)$	$p(z \text{ or } t > \square)$	$2p(z \text{ or } t < \square)$ if $\square$ neg $2p(z \text{ or } t > \square)$ if $\square$ pos

$< \alpha \Rightarrow$ Reject Ho $\quad > \alpha$ Fail to reject.

Commands:

to find z^* invNorm (area to left)

ex. 

invNorm(.95)

qnorm(area to left)

t^* invT (area to left, df)

qT (area to left, df)

---

pvalues: z: normalcdf(low, high)

<: pnorm (□) >: 1-pnorm(□)

t: tcdf (low, high, df)

<: pt (□, df) >: 1-pt(□, df)

$p(t > 2.6) = \text{tcdf}(2.6, 999999, df)$

$1 - \text{pt}(2.6, df)$

$p(t < 2.6) = \text{tcdf}(-99999, 2.6, df)$

$\text{pt}(2.6, df)$

$n - 1 = df$

$n_1 = 34$
$\bar{x}_1 = 3542$

2. A national computer retailer believes that the average sales are greater for salespersons with a college degree. A random sample of 34 salespersons with a degree had an average weekly sale of $3542 last year, while 37 salespersons without a college degree averaged $3301 in weekly sales. The standard deviations were $468 and $642 respectively. Is there evidence to support the retailer's belief?

$n_2 = 37$
$\bar{x}_2 = 3301$

2 sample mean t-test.

$H_0: \mu_1 = \mu_2$

$H_a: \mu_1 > \mu_2$

1 = college degree
2 = no degree

3. Quart cartons of milk should contain at least 32 ounces. A sample of 22 cartons contained the following amounts in ounces. Does sufficient evidence exist to conclude the mean amount of milk in cartons is less than 32 ounces?

31.5, 32.2, 31.9, 31.8, 31.7, 32.1, 31.5, 31.6, 32.4, 31.6, 31.8
32.2, 32.1, 31.8, 31.6, 32.0, 31.6, 31.7, 32.0, 31.9, 31.8, 31.6

one sample t-test

use data to find $\bar{x}$ and S_x.

$H_0: \mu = 32$

$H_a: \mu < 32$

4. *Hippocrates* magazine states that 37 percent of all Americans take multiple vitamins regularly. Suppose a researcher surveyed 750 people to test this claim and found that 290 did regularly take a multiple vitamin. Is this sufficient evidence to conclude that the actual percentage is different from 37%?

One sample prop z test

$H_0: p = .37$

$H_a: p \neq .37$

$$z = \frac{\frac{290}{750} - .37}{\sqrt{\frac{.37(1-.37)}{750}}}$$

5. In a experiment on relaxation techniques, subject's brain signals were measured before and after the relaxation exercises with the following results:

Person	1	2	3	4	5
Before	32	38	65	50	30
After	25	35	56	52	24
* before-after	7	3	9	-2	6

find $\bar{x}$ & S_x for this

Is there sufficient evidence to suggest that the relaxation exercise slowed the brain waves?

matched pairs

$H_0: \mu_D = 0$

$H_a: \mu_D > 0$

$\mu_{\text{before-after}} = 0$

1

6. A private and a public university are located in the same city. For the private university, 1046 alumni were surveyed and 653 said that they attended at least one class reunion. For the public university, 791 out of 1327 sampled alumni claimed they have attended at least one class reunion. Is the difference in the sample proportions statistically significant?

2 sample prop z test

$H_0: p_1 = p_2$

$H_a: p_1 \neq p_2$

$\hat{p}_1 = \frac{653}{1046}$

$\hat{p}_2 = \frac{791}{1327}$

7. Solid fats are more likely to raise blood cholesterol levels than liquid fats. Suppose a nutritionist analyzed the percentage of saturated fat for a sample of 6 brands of stick margarine (solid fat) and for a sample of 6 brands of liquid margarine and obtained the following results:

1 Stick	25.826.926.225.326.726.1	2 sample t-test
2 Liquid	16.917.416.816.217.316.8	

Is there a significant difference in the average amount of saturated fat in solid and liquid fats?

$H_0: \mu_1 = \mu_2$

$H_a: \mu_1 \neq \mu_2$

8. The community hospital is studying its distribution of patients. A random sample of 317 patients presently in the hospital gave the following information:

Type of patient	old rate of occurrences of these types of patients	Exp	present number of occurrences of these types of patients
Maternity ward	20%	.2(317) =	65
Cardiac ward	32%	.32(317) =	100
Burn ward	10%		29
Children's ward	15%		48
All other wards	23%		75
			317

Using a 5% level of significance, test the claim that the distribution of patients in these wards has not changed.

χ^2 goodness of fit

H_0: no difference in present # and old rate

H_a: is a difference

Final Exam outline:

18 m/c (no f/r)

1-13 are 5 pts each and 14-18 are 7 pts each
Topics:

- Probability distributions
- Normal distributions
- Conditional probabilities
- Choosing correct hypothesis test
- Mean and variance (or standard deviation) of probability distributions
- Mena and variance (or standard deviation) of linear combinations of distributions (E[aX+b], Var[aX+b])
- Find c such that $P(Z < c)$ = some value
- Standard error of sample proportion
- Confidence intervals
- Probability rules (several questions on this)
- Binomial distribution probabilities
- Geometric distribution probabilities
- LSRL, r, r^2, residuals
- Hypothesis tests

8) A polling organization announces that the proportion of American voters who favor congressional term limits is 64 percent, with a 95% confidence margin of error of 3 percent. If the opinion poll had announced the margin of error for 80% confidence rather than 95% confidence, this margin of error would be

a) 3%, because the same sample is used.
b) Less than 3%, because we require less confidence.
c) Less than 3%, because the sample size is smaller.
d) Greater than 3%, because we require less confidence.
e) Greater than 3%, because the sample size is smaller.

ME = .03
95% CI

80% → smaller than 95% CI

Popper 24 1-10 A