STAT 400
UIUC

Practice Problems #12

Stepanov
Dalpiaz

SOLUTIONS

The following are a number of practice problems that may be *helpful* for completing the homework, and will likely be **very useful** for studying for exams.

1. **7.1-18** 8.1-18

The April 18, 1994 issue of *Time* magazine reported the results of a telephone poll of 800 adult Americans, 605 nonsmokers, who were asked the following question: "Should the federal tax on cigarettes be raised by $1.25 to pay for health care reform?" Let p_1 and p_2 equal the proportions of nonsmokers and smokers, respectively, who say yes to this question. y_1 = 351 nonsmokers and y_2 = 41 smokers said yes.

$n_1 = 605, \quad y_1 = 351.$ $\qquad \hat{p}_1 = \dfrac{y_1}{n_1} = \dfrac{351}{605} \approx 0.58.$

$n_2 = 195, \quad y_2 = 41.$ $\qquad \hat{p}_2 = \dfrac{y_2}{n_2} = \dfrac{41}{195} \approx 0.21.$

$$\hat{p} = \frac{y_1 + y_2}{n_1 + n_2} = \frac{351 + 41}{605 + 195} = \frac{392}{800} = 0.49.$$

a) With α = 0.05, test $\mathrm{H}_0: p_1 = p_2$ vs. $\mathrm{H}_1: p_1 \neq p_2$.

$\mathrm{H}_0: p_1 = p_2$ vs. $\mathrm{H}_1: p_1 \neq p_2$. Two – tailed.

The test statistic is

$$z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p} \cdot (1 - \hat{p}) \cdot \left(1/n_1 + 1/n_2 \right)}}, \quad \text{where} \quad \hat{p} = \frac{y_1 + y_2}{n_1 + n_2} = \frac{n_1 \cdot \hat{p}_1 + n_2 \cdot \hat{p}_2}{n_1 + n_2}.$$

$\alpha = 0.05$.

The critical (rejection) region is $z < -z_{\alpha/2} = -1.96$ or $z > z_{\alpha/2} = 1.96$.

The observed value of z (test statistic)

$$z = \frac{0.58 - 0.21}{\sqrt{0.49 \cdot 0.51 \cdot \left(1/605 + 1/195 \right)}} = \mathbf{8.988}$$

is greater than 1.96 (the test statistic does fall into the rejection region),

so **Reject $\mathbf{H_0}$**.

b) Find a 95% confidence interval for $p_1 - p_2$.

$$\left(\hat{p}_1 - \hat{p}_2 \right) \pm z_{\alpha/2} \cdot \sqrt{\frac{\hat{p}_1 \cdot \left(1 - \hat{p}_1 \right)}{n_1} + \frac{\hat{p}_2 \cdot \left(1 - \hat{p}_2 \right)}{n_2}}$$

95% confidence level $\alpha = 0.05$ $\alpha/2 = 0.025$. $z_{\alpha/2} = 1.96$.

$$\left(0.58 - 0.21 \right) \pm 1.96 \cdot \sqrt{\frac{0.58 \cdot 0.42}{605} + \frac{0.21 \cdot 0.79}{195}}$$ or **[0.30 , 0.44]**.

c) Find a 95% confidence interval for p, the proportion of adult Americans who would say yes.

$$\hat{p} \pm z_{\alpha/2} \cdot \sqrt{\frac{\hat{p} \cdot \left(1 - \hat{p} \right)}{n}}.$$

95% confidence level $\alpha = 0.05$ $\alpha/2 = 0.025$. $z_{\alpha/2} = 1.96$.

$$\left[0.49 - 1.96 \cdot \sqrt{\frac{0.49 \cdot 0.51}{800}}, 0.49 + 1.96 \cdot \sqrt{\frac{0.49 \cdot 0.51}{800}} \right]$$ = **[0.455 , 0.525]**.

2. A new method of storing snap beans is believed to retain more ascorbic acid than the old method. In an experiment, snap beans were harvested under uniform conditions and frozen in 25 equal-size packages. Ten of those packages were randomly selected and stored according to the new method, and the other 15 packages were stored by the old method. Subsequently, ascorbic acid determinations (in mg/kg) were made, and the following summary statistics were calculated.

	New Method	Old Method
(sample) mean ascorbic acid	435	410
(sample) standard deviation	20	45

a) Use Welch's T to construct a 95% confidence interval for $\mu_{\text{New}} - \mu_{\text{Old}}$.

$$\left\lfloor \frac{\left(\frac{s_x^2}{n} + \frac{s_y^2}{m}\right)^2}{\frac{1}{n-1}\cdot\left(\frac{s_x^2}{n}\right)^2 + \frac{1}{m-1}\cdot\left(\frac{s_y^2}{m}\right)^2} \right\rfloor = \left\lfloor \frac{\left(\frac{20^2}{10} + \frac{45^2}{15}\right)^2}{\frac{1}{10-1}\cdot\left(\frac{20^2}{10}\right)^2 + \frac{1}{15-1}\cdot\left(\frac{45^2}{15}\right)^2} \right\rfloor$$

$$= \lfloor 20.69867 \rfloor = 20 \text{ degrees of freedom.}$$

$t_{0.025}(20) = 2.086$, $\quad (435 - 410) \pm 2.086 \cdot \sqrt{\frac{20^2}{10} + \frac{45^2}{15}}$

$\mathbf{25 \pm 27.6}$ or $\mathbf{(-2.6,\ 52.6)}$.

b) Test $H_0: \mu_{\text{New}} = \mu_{\text{Old}}$ vs. $H_1: \mu_{\text{New}} > \mu_{\text{Old}}$ at a 5% level of significance.

Test Statistic: $T = \frac{(\overline{X} - \overline{Y}) - \delta_0}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} = \frac{(435 - 410) - 0}{\sqrt{\frac{20^2}{10} + \frac{45^2}{15}}} \approx \mathbf{1.890}$.

Critical Value: $t_{0.05}(20) = 1.725$. **Reject H_0 at $\alpha = 0.05$.**

3. Assume that the distributions of X and Y are $N(\mu_1, \sigma^2)$ and $N(\mu_2, \sigma^2)$, respectively. Given the $n = 4$ observations of X,

105, 130, 135, 150

and the $m = 6$ observations of Y,

126, 141, 146, 156, 166, 171

find the p-value (approximately) for the test H_0: $\mu_1 = \mu_2$ vs. H_1: $\mu_1 \neq \mu_2$.

"Hint": Assume the population variances are equal.

$$\bar{x} = \frac{\sum x}{n} = \frac{520}{4} = \mathbf{130}.$$

x	$x - \bar{x}$	$(x - \bar{x})^2$
105	-25	625
130	0	0
135	5	25
150	20	400
	0	1050

$$s_x^2 = \frac{\sum (x - \bar{x})^2}{n-1} = \frac{1050}{3} = \mathbf{350}.$$

$$\bar{y} = \frac{\sum y}{n} = \frac{906}{6} = \mathbf{151}.$$

y	$y - \bar{y}$	$(y - \bar{y})^2$
126	-25	625
141	-10	100
146	-5	25
156	5	25
166	15	225
171	20	400
	0	1400

$$s_y^2 = \frac{\sum (y - \bar{y})^2}{n-1} = \frac{1400}{5} = \mathbf{280}.$$

$$s^2_{\text{pooled}} = \frac{(4-1)\cdot 350 + (6-1)\cdot 280}{4+6-2} = 306.25. \qquad s_{\text{pooled}} = 17.5.$$

Test Statistic: $$T = \frac{(\overline{X} - \overline{Y}) - \delta_0}{s_{\text{pooled}} \cdot \sqrt{\frac{1}{n} + \frac{1}{m}}} = \frac{(130-151) - 0}{17.5 \cdot \sqrt{\frac{1}{4} + \frac{1}{6}}} \approx \mathbf{-1.859}.$$

$n + m - 2 = 4 + 6 - 2 =$ **8** d.f. $\qquad t_{0.05}(8 \text{ d.f.}) = 1.860.$

p-value $(2 - \text{tailed}) \approx 2 \times 0.05 = \mathbf{0.10}$.

4. A random sample of 9 adult elephants had the sample mean weight of 12,240 pounds and the sample standard deviation of 450 pounds. A random sample of 16 adult hippos had the sample mean weight of 5,700 pounds and the sample standard deviation of 400 pounds. Assume that the two populations are approximately normally distributed, Construct a 95% confidence interval for the difference between their overall average weights of adult elephants and adult hippos.

a) Assume that the overall standard deviations are equal.

$$s_{\text{pooled}}^2 = \frac{(9-1)\cdot 450^2 + (16-1)\cdot 400^2}{9+16-2} \approx 174{,}782.6 \qquad s_{\text{pooled}} \approx 418.07$$

$$\left(\overline{X}-\overline{Y}\right) \pm t_{\alpha/2}\cdot s_{\text{pooled}}\cdot\sqrt{\frac{1}{n}+\frac{1}{m}}$$

9 + 16 – 2 = **23** degrees of freedom

$t_{0.025}(23) = 2.069$ $\qquad (12{,}240-5{,}700) \pm 2.069\cdot 418.07\cdot\sqrt{\frac{1}{9}+\frac{1}{16}}$

6,540 ± 360.4 $\qquad$ **(6,179.6 , 6,900.4)**

b) Do NOT assume that the overall standard deviations are equal. Use Welch's T.

$$\left\lfloor \frac{\left(\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}\right)^2}{\frac{1}{n_1-1}\cdot\left(\frac{s_1^2}{n_1}\right)^2+\frac{1}{n_2-1}\cdot\left(\frac{s_2^2}{n_2}\right)^2} \right\rfloor = \left\lfloor \frac{\left(\frac{450^2}{9}+\frac{400^2}{16}\right)^2}{\frac{1}{9-1}\cdot\left(\frac{450^2}{9}\right)^2+\frac{1}{16-1}\cdot\left(\frac{400^2}{16}\right)^2} \right\rfloor$$

$= \lfloor 15.1 \rfloor =$ **15** degrees of freedom

$t_{0.025}(15) = 2.131$ $\qquad (12{,}240-5{,}700) \pm 2.131\cdot\sqrt{\frac{450^2}{9}+\frac{400^2}{16}}$

6,540 ± 384.17 $\qquad$ **(6,155.83 , 6924.17)**

5. Six children are tested for pulse rate before and after watching a violent movie with the following results.

Child	Before	After
1	102	112
2	96	108
3	89	94
4	104	112
5	90	102
6	85	98

Using the paired t test, test for differences in the before and after mean pulse rates.

Use $\alpha = 0.05$ and use a two-sided test.

Before	102	96	89	104	90	85
After	112	108	94	112	102	98
Difference	10	12	5	8	12	13

$$\bar{d} = \frac{10+12+5+8+12+13}{6} = \frac{60}{6} = \mathbf{10}.$$

$\sum d^2 = 100 + 144 + 25 + 64 + 144 + 169 = 646.$

$$s_d^2 = \frac{\sum d^2 - n \cdot \bar{d}^2}{n-1} = \frac{646 - 6 \cdot 10^2}{6-1} = \frac{46}{5} = \mathbf{9.2}.$$

OR

$$s_d^2 = \frac{\sum \left(d - \bar{d}\right)^2}{n-1} = \frac{0+4+25+4+4+9}{5} = \frac{46}{5} = \mathbf{9.2}.$$

$H_0: \mu_B = \mu_A$ vs. $H_1: \mu_B \neq \mu_A$. $\Leftrightarrow$ $H_0: \mu_D = 0$ vs. $H_1: \mu_D \neq 0$.

Test Statistic: $$T = \frac{\bar{d} - 0}{s_d / \sqrt{n}} = \frac{10 - 0}{\sqrt{9.2} / \sqrt{6}} \approx \mathbf{8.076}.$$

$n - 1 = 5$ degrees of freedom.

$\alpha = 0.05$, $\alpha/2 = 0.025$, $\pm t_{\alpha/2} = \pm 2.571$. ← Critical Values

The test statistic **is** in the Rejection Region. **Reject H_0 at $\alpha = 0.05$.**

OR

$n - 1 = 5$ degrees of freedom. $t_{0.005} = 4.032$.

⇒ Right tail is less than 0.005.

⇒ P-value = (two tails) is less than 0.01.

(p-value ≈ 0.000472)

P-value < 0.05 = α. **Reject H_0 at $\alpha = 0.05$.**

OR

Confidence interval: $\bar{d} \pm t_{\alpha/2} \cdot \frac{s_d}{\sqrt{n}}$.

$n - 1 = 5$ degrees of freedom.

$\alpha = 0.05$, $\alpha/2 = 0.025$, $t_{\alpha/2} = 2.571$.

$10 \pm 2.571 \cdot \frac{\sqrt{9.2}}{\sqrt{6}}$ **10 ± 3.2** **(6.8, 13.2)**

95% confidence interval does NOT cover zero ⇔ **Reject H_0 at $\alpha = 0.05$.**

6. Crosses of mice will produce either gray, brown, or albino offspring. Mendel's model predicts that the probability of a gray offspring is $9/16$; the probability of a brown offspring is $3/16$; and the probability of an albino offspring is $4/16$.

a) An experiment to assess the validity of Mendel's theory produces the following data: 35 gray offspring; 20 brown offspring; and 25 albino offspring. Test

$$H_0: p_1 = 9/16,\ p_2 = 3/16,\ p_3 = 4/16 \quad \text{vs} \quad H_1: \text{not } H_0$$

i) at $\alpha = 0.05$, ii) at $\alpha = 0.10$.

	gray	brown	albino
O	35	20	25
E	$80 \cdot 9/16 = 45$	$80 \cdot 3/16 = 15$	$80 \cdot 4/16 = 20$
$\frac{(O-E)^2}{E}$	$\frac{(35-45)^2}{45}$	$\frac{(20-15)^2}{15}$	$\frac{(25-20)^2}{20}$
	2.2222	1.6667	1.25

$$\chi^2 = \sum_{cells} \frac{(O-E)^2}{E} = 2.2222 + 1.6667 + 1.25 = \mathbf{5.1389}. \qquad k - 1 = 3 - 1 = 2 \text{ d.f.}$$

Rejection Region: "Reject H_0 if $\chi^2 \geq \chi^2_\alpha$"

i) $\chi^2_{0.05} = 5.991$. $5.1389 = \chi^2 < \chi^2_\alpha = 5.991$.

Do NOT Reject H_0 at $\alpha = 0.05$.

ii) $\chi^2_{0.10} = 4.605$. $5.1389 = \chi^2 > \chi^2_\alpha = 4.605$.

Reject H_0 at $\alpha = 0.10$.

b) Suppose the experiment in part (a) is repeated, but with twice as many observations. Suppose also that we happened to get the same proportions, namely, 70 gray offspring; 40 brown offspring; and 50 albino offspring. Repeat part (a) in this case, using $\alpha = 0.01$.

	black	brown	albino
O	70	40	50
E	$160 \cdot 9/16 = 90$	$160 \cdot 3/16 = 30$	$160 \cdot 4/16 = 40$
$\frac{(O-E)^2}{E}$	$\frac{(70-90)^2}{90}$	$\frac{(40-30)^2}{30}$	$\frac{(50-40)^2}{40}$
	4.4444	3.3333	2.5

$$\chi^2 = \sum_{cells} \frac{(O-E)^2}{E} = 4.4444 + 3.3333 + 2.50 = \mathbf{10.2778}. \qquad k - 1 = 3 - 1 = 2 \text{ d.f.}$$

Rejection Region: “Reject H_0 if $\chi^2 \geq \chi^2_\alpha$”

$\chi^2_{0.01} = 9.210.$ $\qquad 10.2778 = \chi^2 > \chi^2_\alpha = 9.210.$

Reject H_0 at $\alpha = 0.01$.

7. Suppose we toss a 6-sided die 120 times and count how many times each outcome (1 through 6) occurs. We obtain the following results:

Outcome	1	2	3	4	5	6
Observed frequency	14	24	28	17	12	25

We want to use the chi-square goodness-of-fit test to test the hypothesis that the die is fair (balanced) using a 5% level of significance.

a) State the null hypothesis.

$H_0: p_1 = 1/6,\ p_2 = 1/6,\ p_3 = 1/6,\ p_4 = 1/6,\ p_5 = 1/6,\ p_6 = 1/6$ vs $H_1:$ not H_0

b) Calculate the values of the chi-square test statistic.

O	14	24	28	17	12	25
E	20	20	20	20	20	20
$\frac{(O-E)^2}{E}$	$\frac{(14-20)^2}{20}$	$\frac{(24-20)^2}{20}$	$\frac{(28-20)^2}{20}$	$\frac{(17-20)^2}{20}$	$\frac{(12-20)^2}{20}$	$\frac{(25-20)^2}{20}$
	1.80	0.80	3.20	0.45	3.20	1.25

$$\chi^2 = \sum_{cells} \frac{(O-E)^2}{E} = 1.80 + 0.80 + 3.20 + 0.45 + 3.20 + 1.25 = \mathbf{10.7}.$$

c) Find the critical value χ^2_α.

Rejection Region: "Reject H_0 if $\chi^2 \ge \chi^2_\alpha$"

$k - 1 = 6 - 1 = 5$ d.f. $\chi^2_{0.05} = 11.07.$

d) Test the hypothesis that the die is fair (balanced) using a 5% level of significance.

$10.7 = \chi^2 \not\ge \chi^2_\alpha = 11.07.$ **Do NOT Reject H_0** at $\alpha = 0.05$.

8. An article in *Business Week* reports profits and losses of firms by industry. A random sample of 100 firms is selected, and for each firm in the sample, we record whether the company made money or lost money, and whether or not the firm is a service company. The data are summarized in the 2 × 2 table below. Use a 10% level of significance to test whether the two events "the company made profit this year" and "the company is in the service industry" are independent.

	Industry Type	
	Service	Nonservice
Profit	32	38
Loss	8	22

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

32 **28** *0.571429*	38 **42** *0.380952*	70
8 **12** *1.333333*	22 **18** *0.888889*	30
40	60	100

Q = **3.174603**.

$(2-1)(2-1) = 1$ degree of freedom. $\chi^2_{0.10}(1) =$ **2.706**.

Q = 3.174603 > 2.706 = χ^2_α. **Reject $\mathbf{H_0}$**.

9. A group of 250 children (125 boys and 125 girls) were asked to identify their favorite color. We obtain the following data:

	Favorite Color				
Sex	Red	Green	Blue	Pink	Purple
Boys	35	25	35	10	20
Girls	25	20	25	30	25

A toy manufacturer wants to know if the color preferences of boys and girls differ. Perform χ^2 test of homogeneity using a 1% level of significance.

H_0: In all 5 response categories (Red, Green, Blue, Pink, Purple), the probabilities are equal for these 2 populations (Boys, Girls).

H_A: Not H_0.

O

E

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

	Favorite Color					
Sex	Red	Green	Blue	Pink	Purple	Total
Boys	35	25	35	10	20	125
	30	**22.5**	**30**	**20**	**22.5**	
	0.83333	*0.27778*	*0.83333*	*5.00000*	*0.27778*	
Girls	25	20	25	30	25	125
	30	**22.5**	**30**	**20**	**22.5**	
	0.83333	*0.27778*	*0.83333*	*5.00000*	*0.27778*	
Total	60	45	60	40	45	250

Q = **14.4444**.

$(2-1)(5-1) = 4$ degrees of freedom. $\chi^2_{0.01}(4) = \mathbf{13.28}$.

Q = 14.4444 > 13.28 = χ^2_α. **Reject H_0.**

10. A breakfast cereal manufacturer wants to know whether individual preferences for types of breakfast cereal sweetener are associated with the age of the buyer. In a survey, sweetener preferences were matched to the consumers' age group. From a random sample of 500 responses, the results were as follows:

	Age (Years)			
Cereal	15 – 25	26 – 40	41 – 60	Over 60
Sugar Sweetened	50	60	55	35
Fruit Sweetened	25	35	55	35
Natural	25	55	40	30

Test whether sweetener preferences and age are independent at a 1% level of significance.

	Age (Years)				
Cereal	15 – 25	26 – 40	41 – 60	Over 60	*Total*
Sugar Sweetened	50 **40** *2.5*	60 **60** *0*	55 **60** *0.41667*	35 **40** *0.625*	200
Fruit Sweetened	25 **30** *0.83333*	35 **45** *2.22222*	55 **45** *2.22222*	35 **30** *0.83333*	150
Natural	25 **30** *0.83333*	55 **45** *2.22222*	40 **45** *0.55556*	30 **30** *0*	150
Total	100	150	150	100	500

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

Q = **13.26389**.

$(3-1)(4-1) = 6$ degrees of freedom. $\chi^2_{0.01}(6) = \mathbf{16.81}$.

Q = 13.26389 < 16.81 = $\chi^2_{\alpha}(k-1)$. **Do NOT Reject H_0.**