University Introductory Statistics · Practice Worksheet · tutorinmontreal.ca

Two-Sample Inference and Chi-Square Tests

Name: ______________________________ Date: ______________

Concepts

- Large-Sample Inference for a Difference of Two Means
- The Pooled Two-Sample t Procedure
- The Paired t Test
- Inference for a Difference of Two Proportions
- The Chi-Square Goodness-of-Fit Test
- Contingency Tables and Expected Counts (theory)
- The Chi-Square Test of Independence
- Association Versus Causation in Categorical Data (theory)

Large-Sample Inference for a Difference of Two Means

1. A furniture retailer compares two formats of assembly instructions for the same flat-pack shelf. Of 200 customers chosen at random, 100 receive printed diagrams (format A) and 100 receive a video (format B), independently. The assembly times, in minutes, give

$$\text{A: } n_1 = 100,\ \bar{x}_1 = 42.5,\ s_1 = 8.0 \qquad \text{B: } n_2 = 100,\ \bar{x}_2 = 38.9,\ s_2 = 6.0.$$

Use $z_{0.05} = 1.645$ and $z_{0.025} = 1.960$.

a) Find the standard error of $\bar{x}_1 - \bar{x}_2$.

b) Construct a 95% confidence interval for $\mu_1 - \mu_2$, the difference in mean assembly times.

c) The assembly times are known to be right-skewed. Explain why the interval in (b) is still valid.

2. Two courier companies, X and Y, deliver parcels in the same region. An online retailer claims that Y is faster on average. Random samples of deliveries give the time from order to doorstep, in hours:

$$\text{X: } n_1 = 48,\ \bar{x}_1 = 52.3,\ s_1 = 12 \qquad\qquad \text{Y: } n_2 = 64,\ \bar{x}_2 = 48.0,\ s_2 = 8.$$

Test the retailer's claim at the $\alpha = 0.05$ level, giving hypotheses, conditions, test statistic, decision and a conclusion in context. Report the p-value. Use $z_{0.05} = 1.645$, $z_{0.025} = 1.960$ and

$$P(Z < 2.05) = 0.9798, \quad P(Z < 2.10) = 0.9821, \quad P(Z < 2.15) = 0.9842, \quad P(Z < 2.20) = 0.9861.$$

The Pooled Two-Sample t Procedure

3. A woodworking supplier compares the drying time, in minutes, of two brands of varnish. Independent random samples of test panels give

Brand 1: $n_1 = 10,\ \bar{x}_1 = 48.2,\ s_1 = 3.6$ Brand 2: $n_2 = 12,\ \bar{x}_2 = 44.9,\ s_2 = 4.1.$

Assume both populations of drying times are normal with equal variances. Use $t_{0.025,\,20} = 2.086$, $t_{0.025,\,21} = 2.080$ and $t_{0.025,\,22} = 2.074$.

a) Compute the pooled standard deviation s_p.

b) Construct a 95% confidence interval for $\mu_1 - \mu_2$.

c) Does the interval allow the two brands to have the same mean drying time?

4. A plant nursery grows seedlings of one species under two light regimes. After four weeks, the heights in centimetres of randomly chosen seedlings are summarized as

$$\text{Regime 1: } n_1 = 9,\ \bar{x}_1 = 14.6,\ s_1 = 2.4 \qquad \text{Regime 2: } n_2 = 11,\ \bar{x}_2 = 12.3,\ s_2 = 2.0.$$

Assume heights are normally distributed under each regime with a common variance. Test at $\alpha = 0.05$ whether the mean heights differ, in the five-part form, and bracket the p-value. Upper-tail values for 18 degrees of freedom:

$$t_{0.05,\,18} = 1.734, \quad t_{0.025,\,18} = 2.101, \quad t_{0.01,\,18} = 2.552, \quad t_{0.005,\,18} = 2.878.$$

The Paired t Test

5. A delivery company fits eight of its vans first with standard tyres and then with low-rolling-resistance tyres, and measures each van's fuel consumption in litres per 100 km over the same route:

Van	1	2	3	4	5	6	7	8
Standard	9.8	10.4	11.2	9.6	10.1	10.9	12.0	9.9
Low-resistance	9.2	10.2	10.3	9.2	10.2	10.4	11.3	9.5

Test at $\alpha = 0.05$ whether the low-resistance tyres reduce mean fuel consumption, in the five-part form. Assume the differences are normally distributed. Upper-tail values for 7 degrees of freedom:

$$t_{0.05,\,7} = 1.895, \quad t_{0.025,\,7} = 2.365, \quad t_{0.01,\,7} = 2.998, \quad t_{0.005,\,7} = 3.499.$$

6. Twelve club runners each run a 5 km time trial in their old shoes and, a week later, in a new model. For each runner d = (old time) − (new time) in seconds, and the twelve differences have $\bar{d} = 18.5$ and $s_d = 24.0$. Assume the differences are normal. Use $t_{0.025,\,11} = 2.201$, $t_{0.025,\,12} = 2.179$ and $t_{0.025,\,22} = 2.074$.

a) Construct a 95% confidence interval for μ_d.

b) What does the sign of the interval's endpoints say about the new shoes?

Inference for a Difference of Two Proportions

7. A grocery chain samples 200 customers who shop through its app and, independently, 200 who shop in store. Of the app shoppers, 120 used at least one coupon; of the in-store shoppers, 90 did. Use $z_{0.025} = 1.960$, and take as the condition for the interval at least 10 successes and 10 failures in each sample.

a) Check the condition.

b) Construct a 95% confidence interval for $p_1 - p_2$, the difference between the proportions of app and in-store shoppers who use a coupon.

8. A dental clinic compares two appointment reminders. Of 160 patients chosen at random to receive a text message, 48 missed their appointment; of 150 patients chosen at random to receive a phone call, 30 missed theirs. Test at $\alpha = 0.05$ whether the no-show rates differ, in the five-part form, and give the p-value. Take as the condition at least 10 expected successes and failures in each sample under H_0. Use $z_{0.025} = 1.960$ and

$$P(Z < 2.00) = 0.9772, \quad P(Z < 2.03) = 0.9788, \quad P(Z < 2.05) = 0.9798.$$

The Chi-Square Goodness-of-Fit Test

9. A municipal helpline wants to know whether calls are spread evenly across the five weekdays. In a random sample of 200 calls, the counts from Monday to Friday are $32,\ 38,\ 41,\ 47,\ 42$. Test at $\alpha = 0.05$ whether the calls are uniformly distributed over the weekdays, in the five-part form; the test requires every expected count to be at least 5. Use $\chi^2_{0.10,4} = 7.779$, $\chi^2_{0.05,4} = 9.488$ and $\chi^2_{0.05,5} = 11.070$.

10. A ferry operator's planning assumes that its passengers are 60% foot passengers, 25% travelling with a car and 15% with a bicycle. A random sample of 160 passengers contains 80 on foot, 52 with a car and 28 with a bicycle. Test the assumption at $\alpha = 0.05$ in the five-part form and bracket the p-value. The test requires every expected count to be at least 5. Upper-tail values:

$$\chi^2_{0.05,\,2} = 5.991, \quad \chi^2_{0.025,\,2} = 7.378, \quad \chi^2_{0.01,\,2} = 9.210, \quad \chi^2_{0.05,\,3} = 7.815.$$

Contingency Tables and Expected Counts

11. A random sample of 200 students at two campuses of one university is classified by campus and by usual way of getting to class:

	Car	Public transit	Walk or cycle	Total
Downtown campus	46	48	26	120
Suburban campus	44	22	14	80
Total	90	70	40	200

a) Find the expected count of every cell under the hypothesis that campus and mode of travel are independent.

b) Check the condition that every expected count is at least 5, and give the degrees of freedom of the chi-square test of independence for this table.

The Chi-Square Test of Independence

12. A random sample of 150 adults is asked whether they drink at least two cups of coffee a day and whether they usually sleep fewer than 7 hours a night:

	Under 7 h	7 h or more	Total
Two or more cups	42	38	80
Fewer than two	18	52	70
Total	60	90	150

Compute the chi-square statistic for independence, state its degrees of freedom, and decide at $\alpha = 0.05$. The test requires every expected count to be at least 5. Use $\chi^2_{0.05,\,1} = 3.841$, $\chi^2_{0.005,\,1} = 7.879$ and $\chi^2_{0.05,\,2} = 5.991$.

13. A random sample of 240 undergraduates, 120 in science programmes and 120 in arts programmes, is asked where they usually study:

	Library	Home	Café	Total
Science	54	38	28	120
Arts	36	52	32	120
Total	90	90	60	240

Test at $\alpha = 0.05$ whether study location is independent of programme, in the five-part form, and bracket the p-value. The test requires every expected count to be at least 5. Upper-tail values:

$$\chi^2_{0.05,\,2} = 5.991, \quad \chi^2_{0.025,\,2} = 7.378, \quad \chi^2_{0.05,\,3} = 7.815, \quad \chi^2_{0.05,\,5} = 11.070.$$

Association Versus Causation in Categorical Data

14. A random sample of households in one city is classified by whether the home has a smoke detector that works and whether the household reported a kitchen fire in the past five years. A chi-square test of independence gives $p = 0.002$. For each statement, say whether the study supports it, and give a one-line reason.

a) Having a working smoke detector and reporting a kitchen fire are associated in this city's households.

b) Installing a working smoke detector changes a household's chance of a kitchen fire.

c) If the two variables were independent, a table at least this far from the expected counts would occur in about 0.2% of samples.

d) The result applies to households in every city.

Synthesis — drawing on several topics in this unit

15. A library tests two wordings of an overdue-book email. Of 120 borrowers chosen at random to receive wording 1, 45 return the book within a week; of 120 receiving wording 2, 30 do. The z test requires at least 10 expected successes and 10 expected failures in each group under H_0, and the chi-square test every expected count at least 5. Use $\chi^2_{0.05,\,1} = 3.841$, $z_{0.025} = 1.960$ and $P(Z < 2.09) = 0.9817$.

a) Carry out the two-sided pooled z test of $H_0\colon p_1 = p_2$ and give its p-value.

b) Arrange the data as a 2×2 table and compute the chi-square statistic for independence of wording and prompt return.

c) Compare z^2 with χ^2, and compare the two decisions at $\alpha = 0.05$.