

Hypothesis Testing and Linear Regression.

1. The manufacturer of a patient medicine claimed that it was 90% effective in relieving an allergy for a period of 8 hours. In a sample of 200 people suffering from the allergy, the medicine provided relief for 160 people.

Determine whether the manufacturer’s claim is legitimate. Is your test one- or two-tailed?

Type of test: One Sample Test for Proportions

2. A manufacturer claims that through the use of a fuel additive, automobiles should achieve on average an additional 3 miles per gallon of gas. A random sample of 100 automobiles was used to evaluate this product. The sample mean increase in miles per gallon achieved was 2.4 and the sample standard deviation was 1.8 miles per gallon.

Test the null hypothesis that the population mean is at least 3 miles per gallon. Find a p -value of this test, and interpret your findings. Clearly state all the assumptions you made.

Type of test: One Sample Test for Mean (t-test / Z-approximation)

3. To determine the effectiveness of a feed supplement for cattle, High Grade Feeds conducts a pilot study. Eight steers are randomly divided into groups of 4 steers to be fed with supplement, and another group of 4 steers to be fed exactly the same diet except for the supplement. The weight gains (lbs) over a period of 6 months were as follows:

Supplement Group	330	360	400	350
Control group	290	320	340	370

- (a) Construct a 95% confidence interval for the effect of the supplement.
- (b) Suppose steers wholesale for 90 cents per pound. If the supplement costs \$80 per steer for the whole 6-month period, is it profitable to use it?

Type of test: Two-Sample Test for Means (Case III: Welch t-Test)

4. A random sample of 3864 first marriages showed 5-year survival characteristics, by wife’s church attendance (simplified version of *Balakrishnan et al*, 1987):

	Church Attendance			Totals
	Weekly or More	Sometimes	Rarely or Never	
Still Married	1071	1339	1082	3492
Divorced or Separated	42	128	202	372
Totals	1113	1467	1284	3864

- (a) Calculate a 95% confidence interval for the 5-year survival rate of all marriages. Then repeat three times for the three subpopulations defined by the three levels of church attendance.
- (b) Calculate a 95% confidence interval for the difference in survival rates – for marriages where the wife attends church “weekly or more” vs “rarely or never”.
- (c) How well does the data support the conclusion that regular church attendance helps marriages survive?

Type of test: Confidence Intervals for Proportions and Two-Sample Test for Proportions

5. An investigator wants to compare two population proportions and intends to take independent random samples of the same size from each population. She wants to be sure that a 90% confidence interval for the difference between the two population proportions extends no further than 0.05 on each side of the difference between the sample proportions. How large a sample should she take from each population?

Type of test: Sample Size Determination for Two-Sample Proportions

6. A special computer-based training program designed to increase reading speed was tested on a group of persons. Their results before and after the program are (in minutes):

Before	20	25	23	18	24	25	30	20	19	24	24
After	20	20	21	19	21	22	28	18	18	24	27

- (a) Test at the 5% significance level the hypothesis that the program significantly increases the speed of reading.

- (b) State any assumptions you made in a).
- (c) Give a 95% confidence interval for the difference in means.
- (d) On the basis of the data alone, would you consider trying this program if you wanted to improve your speed of reading? Provide an explanation with your answer.

Type of test: Paired t-Test

7. A study was conducted to determine the amount of hours spent on Facebook by university and high school students. For this reason, a questionnaire was administered to a random sample of 16 university and 14 high school students, and the hours per day spent on Facebook were recorded. Summaries of the data are shown in the table below:

	Sample size	Sample mean	Sample variance
University students	16	2.9	0.9
High school students	14	2.1	1.1

- (a) Use an appropriate hypothesis test to determine whether the mean hours per day spent on Facebook were different between university and high school students. Test at two suitable significance levels, stating clearly the hypotheses, the test statistic, and its distribution under the null hypothesis. Comment on your findings.
- (b) State clearly any assumptions you made in a).
- (c) Adjust the procedure above to determine whether the mean hours spent per day on Facebook for university students is higher than that of high school students.

Type of test: Two-Sample Test for Means (Case III: Welch t-Test) and optionally F-Test for Equality of Variances

8. (8 pts) Health advisory services state that 50% of heart attack patients die before reaching the hospital; the chances of surviving a heart attack are dramatically increased by receiving medical help as quickly as possible. A survey on 240 heart attack victims discovered that men sought help more urgently than women.

	Number	Mean and standard deviation to reach hospital after the initial attack
Men	134	$\bar{x} = 3.40, s = 1.42$
Women	106	$\bar{x} = 8.24, s = 4.24$

- (a) Calculate the 90% confidence interval for the difference between the mean times to reach the hospital for men and women.
- (b) State clearly any assumptions you made in a).
- (c) Is it possible to infer that there is no difference between the mean times taken to reach the hospital from the data given above?

Type of test: Two-Sample Test for Means (Case III: Variances Unknown and Unequal, Welch t-Test)

9. A researcher is interested in determining whether a particular pill provides effective treatment for stomach pain. A randomised experiment was conducted to address this question. The study randomly allocated 200 people to either a group where the pill was administered, or a group where a placebo pill was given. These people were monitored and the numbers of those who got better (or did not) were recorded. The results are summarised below:

	Did not get better	Got better
Pill	21	74
Placebo	50	55

- (a) Give a 95% confidence interval for the difference in the probabilities of getting better from stomach pain between those who took the pill and the placebo group.
- (b) Carry out an appropriate hypothesis test at the 5% significance level to determine whether the probability of getting better is higher in the pill group, compared to the probability in the placebo group. State the test hypotheses, and specify your test statistic and its distribution under the null hypothesis. Comment on your findings.
- (c) State any assumptions you made in part 8b.
- (d) On the basis of the data alone, would you conclude that the particular pill increases the chances of getting better from stomach pain? Provide an explanation with your answer.

Type of test: Two Sample Test for Proportions

10. Plastic sheets produced by a machine are periodically monitored for possible fluctuations in thickness. If the true variance in thickness exceeds 2.25 square millimeters, there is a cause for concern about product quality. Thickness measurements for a random sample of ten sheets produced in a particular shift were taken, giving the following results (in millimeters):

226 226 232 227 225 228 225 228 229 230

- (a) Find the sample variance.
(b) Test at the 5% significance level the null hypothesis that the population variance is at most 2.25.

Type of test: One Sample Test for Variance (Chi-square test for a population variance)

11. A pet food company, **Paws and Reflect**, claims there is **no association** between type of pet ownership and food preference. To test this, they surveyed 300 customers and recorded the following observed frequencies:

Food Preference	Dog Owners	Cat Owners	No Pets
Organic Food	45	30	25
Non-Organic Food	30	40	35
No Preference	25	30	40

- (a) State the null and alternative hypotheses.
(b) Calculate the expected frequencies for all cells.
(c) Compute the chi-squared test statistic.
(d) Determine the degrees of freedom, critical value at $\alpha = 0.05$, and conclude whether to reject H_0 .

12. A researcher collected data on two variables, x and y , for 12 observations. The data is provided below:

x	2	3	4	5	6	7	8	9	10	11	12	13
y	4	5	6	8	7	9	10	11	12	13	14	15

- (a) Estimate the Ordinary Least Squares (OLS) linear regression line $\hat{y} = b_0 + b_1x$. Show all steps, including sums of x , y , xy , and x^2 . Use a calculator for computations.
(b) Interpret the slope coefficient b_1 in the context of the data.