MATH1712 Probability and Statistics II
Homework 8

http://www1.maths.leeds.ac.uk/~voss/2017/MATH1712/
Jochen Voss, J.Voss@leeds.ac.uk
2017/18, semester 2

This exercise sheet will be discussed in the tutorials of the week beginning 23rd April.

For discussion during the tutorial. The main topic of this exercise sheet are hypothesis tests.

(1) The p-value of a statistical test is defined as the probability of obtaining a value of the test statistic which is more extreme than (or the same as) the observed value, assuming H_0 is true. For example, if we have $z = 1.834$ in a two-sided z-test, the p-value is

$$p = P\bigl(|Z| > 1.834\bigr) = 2\Phi(-1.834) = 0.0667.$$

For which p-values can we reject H_0? Why does the statement "H_0 is true with probability p" make no sense? If H_0 is true, what is a typical p-value?

(2) In lectures we have discussed how we can test whether variates observed for two *different* populations have the same mean. Assume that we have observed two variates for each individual of a population instead, *i.e.* we have observed paired samples $(x_1, y_1), \ldots, (x_n, y_n)$. Find examples of observations of this type. How can we test whether x and y have the mean in this situation?

(3) Try to solve the following exam question from the 2014 MATH1725 paper:

a) The absenteeism rates in days and parts of days for nine employees of a large company were recorded in two consecutive years:

employee	1	2	3	4	5	6	7	8	9	total
year 1	3.0	6.7	11.3	5.0	9.4	15.7	8.0	10.0	9.7	78.8
year 2	2.8	5.1	8.4	5.0	6.2	12.2	10.0	6.8	6.0	62.5

Construct a hypothesis test to determine, if there is any evidence that the absenteeism rate is different for the two years. Carry out your test.

b) Obtain a 95% confidence interval for the mean absenteeism rate in year 2. [We may not yet have covered this by the time of the tutorial.]

c) State all assumptions you have used in your answer to part (a) above.

d) Suppose now that nine results for year 1 for one groups of employees and nine results for year 2 for a completely different groups of employees are made available. Outline your method of analysis for determining whether the average absenteeism rate is different for the two years in this case. Give all the appropriate equations but do not do any numerical calculations.

Homework questions. Your solutions to these questions contribute towards your final mark for the module. Each question requires a single letter answer. Submit your answers online at

`https://goo.gl/forms/LCYaPV6Dox6JsRxx2`

before the deadline of **Friday, 4th May, 5pm**.

Exercise 25. For n observations $x_1, \ldots, x_n$, the sample median is m, and the lower and upper quartiles are q_1 and q_3, respectively. Which of the following quantities can be used as a measure of spread?

A: $\dfrac{m}{2}$ B: $\dfrac{q_3 - q_1}{2}$ C: $\dfrac{q_1 + q_3}{2}$ D: $\dfrac{q_1 q_3}{2}$

Exercise 26. Consider the following statements about the sample correlation r_{xy}.

(i) Depends on the units of measurement of X and Y.
(ii) Can never *exactly* equal zero.
(iii) Measures association between variables X and Y.
(iv) Lies between −1 and +1.

Which of these statements are true?

A: (i), (ii), and (iii) B: (i), (iii), and (iv) C: (ii) and (iii) D: (ii) and (iv) E: (iii) and (iv)

Exercise 27. For n observations $x_1, x_2, \ldots, x_n$ the sample mean is $\bar{x}$. Consider the following statistics:

(i) $\frac{1}{n-1}\sum_{i=1}^{n} x_i^2 - \frac{n}{n-1}\bar{x}^2$
(ii) $\frac{1}{n-1}\sum_{i=1}^{n} x_i^2 + \frac{n-1}{n}\bar{x}^2$
(iii) $\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$
(iv) $\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2$

Which of these statistics equal the sample variance?

A: (i), (ii), and (iii) B: (i) and (ii) C: (i) and (iii) D: (i) and (iv) E: (ii) only

Exercise 28. Let $X_1, \ldots, X_n$ be independent and identically distributed random variables, with $\mathbb{E}(X_i) = \mu$ and $\mathrm{Var}(X_i) = \sigma^2$. Define $Y = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} X_i$. Consider the following statements:

(i) $\mathrm{Var}(Y) = \frac{\sigma^2}{n}$
(ii) $\mathbb{E}(Y) = \sqrt{n}\mu$
(iii) $\mathrm{Var}(Y) = \sigma^2$
(iv) $\mathrm{Var}(Y) = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$, where $\bar{X}$ is the sample mean of the X_i

Which of these statements are true?

A: (i), (ii), and (iv) B: (i) and (iii) C: (ii), (iii), and (iv) D: (ii) and (iii) E: (iii) and (iv)

Exercise 29. What distribution could be used to approximate a t-distribution with $\nu = 99$ degrees of freedom?

A: $\chi^2(99)$ B: $B(99, 1/2)$ C: Pois(99) D: $\mathcal{N}(0, 1)$

Exercise 30. Let X and Y be random variables with $\mathbb{E}(X) = 1$, $\mathrm{Var}(X) = 4$, $\mathbb{E}(Y) = 2$, and $\mathrm{Var}(Y) = 9$. Assume that the correlation between X and Y is $\mathrm{Corr}(X, Y) = 1/3$. What is the variance of $3X - 2Y + 1$?

A: 24 B: 36 C: 48 D: 60 E: 72

Exercise 31. Least squares regression is used to fit a regression line $y = \alpha + \beta x$ to data (x_i, y_i) for $i = 1, 2, \ldots, n$. Consider the following statements.

(i) $\hat{\alpha}$ and $\hat{\beta}$ are chosen to minimise $\sum_{i=1}^{n}(y_i - \beta x_i)^2$.

(ii) $\hat{\alpha} = \bar{y} - \hat{\beta}\bar{x}$, where $\bar{x}$ and $\bar{y}$ are the sample means.

(iii) $\hat{\beta} = \mathrm{s}_{xy}/\mathrm{s}_x^2$, where s_{xy} is the sample covariance and s_x^2 is the sample variance.

(iv) $\hat{\beta} = \mathrm{r}_{xy}/\mathrm{s}_x$, where r_{xy} is the sample correlation and s_x is the sample standard deviation.

Which of these statements are true?

A: (i), (ii), and (iii) B: (i) and (iii) C: (ii) and (iii) D: (ii) and (iv) E: (iii) and (iv)

Exercise 32. In testing a null hypothesis H_0 against an alternative H_1, what is a type I error?

A: reject H_0 when H_0 is false

B: accept H_0 when H_0 is true

C: accept H_0 when H_0 is false

D: reject H_0 when H_0 is true

Exercise 33. If a random variable X has a $\chi^2(5)$-distribution, for what value of q is $P(X \leq q) = 0.95$?

A: 1.960 B: 7.779 C: 9.236 D: 9.488 E: 11.070

Exercise 34. A sample of size $n = 25$ from a normal distribution with known variance $\sigma^2 = 1$ is used to construct a 95%-confidence interval for the unknown mean μ. What is the width of the confidence interval?

A: 0.119 B: 0.392 C: 0.415 D: 0.516 E: 0.784