Semester 1 2017,
Example Exam 2

Instructions

The exam consists of 4 questions, 1-4. Each question has four items, a-d.

Within each question:

Item (a) carries a weight of 8 marks.

Item (b) carries a weight of 7 marks.

Item (c) carries a weight of 6 marks.

Item (d) carries a weight of 4 marks.

The total marks in the exam are 100.

Answer ALL questions in the spaces provided.

If more space is required, use the back of the PREVIOUS page.

Show all your working and include sketches where appropriate.

Work written in the formulae and tables booklet will NOT be marked.

Question 1:

The weight of a component (in kilograms) is continuous random variable, X with support $[0,1]$ and pdf

$$f(x) = \begin{cases} Ke^{-x}, & x \in [0,1], \\ 0, & \text{otherwise}, \end{cases}$$

where K is some constant.

(a) Find K.

(b) Calculate an expression for the cdf of X and plot it.

(c) Assume now that 7 such independent random variables are measured and let N denote the number of those components having weight greater than $1/2$. What is $P(N \geq 2)$?

(d) Continuing on the previous item, for each of the 7 independent random variables, you are paid \$5 if it has a weight less than $1/2$ and \$3 otherwise. What is the variance of your total earnings.

Question 2:

A homework assignment consists of 10 questions. The chance of succeeding in a question is $p = 0.8$ and the success/failure of each question is independent of the others.

There are two marking schemes:

Scheme A: All questions are marked and each is worth 10%.

Scheme B: Only a randomly selected subset of size 5 is marked and each is worth 20%.

In both schemes a question is either correct (success) or wrong (failure). Let X denote the percent grade under scheme A and let Y denote the percent grade under scheme B.

(a) What is the mean of X? How about the mean of Y? In this respect which marking scheme is better? Or are the schemes equivalent?

(b) What is $P(X = 100)$? How about $P(Y = 100)$? Now in this respect which marking scheme is better for the student?

(c) What is the variance of X? How about the variance of Y? In this respect, which marking scheme is better? In what way is it better?

(d) Failure occurs if the percent grade is less than 50%. Compare the chance of failure in both schemes. Which scheme is better for the student?

Question 3:

You observe the following output:

```
In [5]: UnequalVarianceTTest(data1,data2)

Out[5]: Two sample t-test (unequal variance)
        ------------------------------------
        Population details:
            parameter of interest:   Mean difference
            value under h_0:         0
            point estimate:          -1.4516996384754872
            95% confidence interval: (-5.91045104330129,3.0070517663503153)

        Test summary:
            outcome with 95% confidence: fail to reject h_0
            two-sided p-value:           0.47361548935859044 (not significant)

        Details:
            number of observations:   [7,8]
            t-statistic:              -0.7524144182475475
            degrees of freedom:       7.902492459938804
            empirical standard error: 1.9293883839395962
```

(a) Present a 99% confidence interval for the difference in means under the assumption of unequal variances.

(b) You now know that the sample standard deviation of "data2" is 2.028. What is the sample standard deviation of "data1"?

(c) Use the same data to carry out a one-sided hypothesis test (still assuming unequal variances): $H_0 : \mu_1 = \mu_2 + 1.2$ vs. $H_1 : \mu_1 > \mu_2 + 1.2$ with $\alpha = 0.01$ and rounding the degrees of freedom up.

(d) Carry out the same test using the assumption of equal variances.

Question 4:

Assume the hypothetical situation where you have observations:

$$
\begin{aligned}
(x_1, y_1) &= (10, 10), \\
(x_2, y_2) &= (20, 22), \\
(x_3, y_3) &= (30, 32), \\
(x_4, y_4) &= (40, 40).
\end{aligned}
$$

(a) Calculate the sample mean and sample variance for both (x_1, x_2, x_3, x_4) and (y_1, y_2, y_3, y_4).

(b) Compute (by hand) the least squares estimates, for the model,

$$Y_i = \beta_0 + \beta_1 x_i + \epsilon_i, \qquad i = 1, 2, 3, 4,$$

where $\epsilon_i \sim N(0, \sigma^2)$.

(c) Sketch a plot of the regression line on a plane containing the three data points. Explain why this line minimises the sum of squares agrees with your answer to (b).

(d) Assume now that after estimating the parameters, you use the model,

$$Y = \hat{\beta}_0 + \hat{\beta}_1 x + \epsilon,$$

where $\hat{\beta}_0$ and $\hat{\beta}_1$ are according to your answer in (b) but ϵ is **uniformly** distributed (as opposed to Normally distributed) with mean 0 and variance 1. Sketch the cdf of Y under this model when $x = 20$.