

Solutions to the exam in HMET5140 fall 2018.

Exercise 1. a) There are 1946 patients in the old treatment arm and 1163 patients in the new treatment arm (1p). The average difference is 1053 EUR (1p), and the new treatment is the most expensive (1p). This is significant at 5%-level (two-sided $p=0.03$) (1p), but not at 1% level (1p). The confidence interval for the difference is from 104 to 2001 EUR (1p), and it is significant, as it does not include 0 (1p).

b) A standard two-sample t-test assumes equal variances in the two groups. The variance ratio test is a test on whether this holds or not (2p). In this case, it does not hold ($p<0.001$), and the unequal variance option has to be used in the t-test (1p).

c) This is a Wilcoxon/Mann-Whitney test, based on comparing rankings instead of actual observations. This removes influence of extreme observations, which could be a problem when comparing means (although the sample size here is large). As such it is more robust towards outliers (and also reduces variation compared with original data) compared with the t-test (2p). The conclusion is somewhat similar in that the costs are significantly different, but with a smaller p-value (<0.001) (1p). A one-way ANOVA (parametric) or Kruskal-Wallis test (non-parametric) could have been used for three groups (1p).

d) On the original scale, the coefficient becomes $\exp(0.377)=1.46$ (1p) with 95% confidence interval= $(\exp(0.316)=1.37, \exp(0.439)=1.55)$ (1p). This means that the (geometric mean) costs are 46% higher in the new treatment group, or from 37% to 55% higher (1p). As the interval does not contain 1, this is significant as we now have a ratio-type measure (1p).

e) The age group 70-79 years is associated with the highest costs. Patients aged 80 years and up is the reference, hence the 70-79 years have $\exp(0.092)=1.096=9.6\%$ higher mean costs than these. All the other categories have either less higher costs than the reference, or lower costs than the reference (negative coefficients) (2p). Patients aged 18-49 have $\exp(-0.126)=0.88$ or 12% lower costs than 80 and up (1p), if you have spent 60 days in hospital, it amounts to $\exp(0.417*2)=2.30$ times higher costs, since lospy is coded per 30 days (1p). The fact that the treatment coefficients are similar could imply that the two treatment arms are well balanced on gender, age and comorbidities (at least when considering costs), which is always a good feature of a study (3p), but most of the background variables have little impact on the costs.

f) Here the coefficients are already exponentiated, hence the treatment coefficient implies that patients receiving the new treatment have 9.4% higher mean costs (2p) (side note: the gamma GLM operates on arithmetic mean scale, which is nicer in terms of interpretation than the manual log-linear model). There is no longer a significant difference between arms at 5% level, although borderline (1p). This is a lower effect than in the log-linear model, and could indicate differences in model fit (1p).

Another side note: If you wonder about the large differences between the log-linear and gamma model, it is not only due to different model fit. The gamma model coefficients actually compare arithmetic mean costs for different groups, whereas the log-linear model compares geometric mean costs. This is of course an advantage of the gamma model here, as it is more straightforward to interpret. In these data, the two turn out to be quite different, and it seems easier to get significant results for geometric mean coefficients.

g) A marginal effect express the effect of a single independent variable from a possibly adjusted regression model, when all other variables are kept at their observed values in the data set. It is

based on predictions from the model: What would the average cost have been if all patients belonged to the new treatment arm, and correspondingly if all belonged to the old treatment arm (3p). As the model in f) is used as a basis for the marginal effects, it means that one gets slightly different means than in a) as the cost difference between treatment arms are now adjusted for age, gender and comorbidities (2p).

h) $H_0: p_{\text{dead new treatment}} = p_{\text{dead old treatment}}$ $H_1: p_{\text{dead new treatment}} \neq p_{\text{dead old treatment}}$ (2p)

Expected cell counts:

d365	treatment:	0	1
0		$2306 \cdot 1946 / 3109 = 1443$	$2306 \cdot 1163 / 3109 = 863$
1		$803 \cdot 1946 / 3109 = 503$	$803 \cdot 1163 / 3109 = 300$

(2p)

$$\text{Test statistic: } \frac{(1219-1443)^2}{1443} + \frac{(1087-863)^2}{863} + \frac{(727-503)^2}{503} + \frac{(76-300)^2}{300} = 361 \quad (2p)$$

This should be compared to the 95%-percentile of a chi-square distribution with 1 d.f., which is 3.84. The conclusion is then to reject H_0 (2p). If one calculates the % dead in the two arms, it becomes 37% in old treatment arm vs 6.5% in new treatment arm! Hence, not surprising that test is significant.

i) Most patients do not die within a year, regardless of treatment arm. Hence, the histogram will have a high column at 1, which is equal to the maximum value of the life year variable. Further, the data could either be skewed to the left, indicating a square or cubic transformation, or if most who die early on, there could be additional high columns near 0 in the histogram, which it would be hard to fix with transformations (3p). A standard t-test would be problematic in any case, as the variable would violate the normality assumption (1p).

j) The Spearman coefficient indicates a negative relationship between costs and life years: higher costs are associated with lower survival (1p). However, it is not significant, with a p-value of 18.8% (1p). As Spearman uses rankings, the fact that neither costs nor life years are normal is not a problem. However, the scatter plot indicates a non-monotonous relationship (violating an assumption of the Spearman coefficient) between the variables, and based on the pattern it does not seem very meaningful to consider a correlation measure like this as the life year variable is not *that* continuous (2p).

k) The output means that the extra cost of new treatment is 4217 EUR per additional life year gained. The output shows that costs and life years are higher in the new treatment group, hence this will be the interpretation of the ICER (in addition to the fact that it was defined like this previously) (2p). Whether it is cost-effective depends on the willingness-to-pay. However, thresholds for price per additional life years is usually much higher than this, so it should be an indication of cost-effectiveness (2p).

l) The bootstrapping is done as follows: Each bootstrap sample is constructed by sampling with replacement from the data set 3109 times (side note: here one could consider stratified sampling to get exactly 1946 and 1163 patients in each arm in every bootstrap sample, but as this is a randomized study were it is hard to guarantee in advance that one gets exactly the number of patients in each arm as was observed, the number of patients in each arm can also be allowed to vary between bootstrap samples). Then, an estimate for the ICER is calculated and stored. This is repeated 10000 times. One then sort the 10000 ICER estimates and select observation no. 250 and 9750 as the lower and upper boundaries of the confidence interval (4p). The histogram shows that

the bootstrap distribution is symmetric, so there should not be a problem with the percentile interval (2p).

Grading:

0-24p=F

25-29p=E

29-35p=D

36-43p=C

44-53p=B

54-60p=A