**Exam HMET5140 2022 solutions.**

a) There are 40 patients in regimen A and 68 patients in regimen B (1p). The lower numbers in each cell shows the percentage of patients with costs over/under 1000 euros within the treatment regimens (1p).

Null hypothesis: $p_{\text{cost>1000 euro, regimen A}} = p_{\text{cost>1000 euro, regimen B}}$

Alternative: $p_{\text{cost>1000 euro, regimen A}} \neq p_{\text{cost>1000 euro, regimen B}}$

or that the probability of having a cost over 1000 Euros is the same for both treatments, whereas the alternative is that there is a difference (1p).

Expected frequencies:

Row 1, col 1: 63*40/108=23.3

Row 1, col 2: 63*68/108=39.7

Row 2, col 1: 45*40/108=16.7

Row 2, col 2: 45*68/108=28.3 (1p)

Test statistic: $\frac{(18-23.3)^2}{23.3} + \frac{(45-39.7)^2}{39.7} + \frac{(22-16.7)^2}{16.7} + \frac{(23-28.3)^2}{28.3} = 4.6$ (1p)

This should be compared to the 95%-percentile of a chi-square distribution with 1 d.f, which is 3.84 from tables of the chi-square distribution (1p). Reject the null hypothesis since 4.6>3.84 (1p).

b) The sdtest is to check whether the variation in the cost distributions can be assumed equal or not, which is used for selecting the appropriate t-test in the next step (1p). The conclusion is that the variation can be assumed equal, with p-value 47% (1p). The null hypothesis for the t-test is that the mean costs are equal, and the alternative is that the mean costs are different (1p). The conclusion is that the mean costs seem to be equal with a p-value of 7.9%, but only just using a 5% significance level (2p).

c) The non-parametric test is a Wilcoxon rank sum/Mann-Whitney test (1p). The null hypothesis is that the distribution of costs (or roughly the median) are equal across the treatment regimens, and the alternative is that the distributions are different (1p). The conclusion here is to reject the null hypothesis, with a p-value of 1.9% (1p). The test does not use the actual observed cost values for each individual, but rather rankings of the observations from lowest to highest value over both groups combined, then summing the rankings for each group and doing a test on whether the rank sums are different or not (2p).

d) Stata samples with replacement 10000 times from the data, not stratified by treatment hence the number of observations in each treatment regimen will vary across samples. For each sample, the difference in means are calculated and stored (in file boot2). They are shorted from lowest to highest value, and observation no. 250 and 9750 will be the lower and upper limits of the 95% confidence interval (2p). The conclusion is still similar to the standard t-test in b), though, as the confidence interval just contains 0. Hence, no difference in costs (1p). The histogram tells us that the bootstrap estimates for mean difference are approximately normal even though the raw data were not - hence a percentile type interval should be ok (2p).

e) The log-transformation should make the cost data more closely normally distributed, as the costs are all positive with a long right tail (1p). The conclusion is the treatment once more is significant,

with p-value = 1.8% (1p). The interpretation is that the costs of treatment B are exp(-0.80)=0.45=55% lower for regimen B compared to regimen A (2p). The 95% confidence interval is calculated as (exp(-1.46), exp(-0.14))=(0.23, 0.87) (1p). Usually, non-parametric tests or log-transforms reduce effects of extreme observations in the upper tail compared to a standard t-test. If this was the case here, one would expect a significant result from the t-tests and non-significant results from the non-parametric/log-transformed test. But here, the opposite is the case. This could be due to the fact that both rankings and log-transformed values has relatively less variation than the raw cost values. The histograms in c) showed huge variation in both groups. Hence, one actually gets greater statistical power in the non-parametric/log-transformed tests compared with the t-tests, yielding p-values going from just above 5% in t-tests to below 5% in the non-parametric/log-transformed tests (2p).

f) As men are coded as 0, the gender coefficient means that women have exp(0.41)=1.51=51% higher costs than men, but this is not significant (2p). The effect of increasing age by 12 years on the log-costs is 0.059*12=0.71 (1p), corresponding to exp(0.059*12)=2.02=more than twice as high costs (1p). The treatment coefficient in e) is similar to the one in f). This means that the age and gender distribution of the patients are similar in both treatment groups (2p).

g) The p-values for the coefficients are all similar, although gender is a bit closer to being significant at the 5% level. However, the numerical values for the anti-logged effects of the independent variables still seem fairly similar, e.g. 1.49 vs 1.51 for gender. The effect of treatment in the log-linear model seems slightly larger (lower costs for regimen B), but with similar p-value. Hence it does not seem to matter much which model is used (3p). Comparing the AIC values, where smaller value indicates better fit, the AIC value for the log gamma model (1717) is slightly lower than the one from the log-linear model (1736). Hence, there are some indications that the log-gamma model fits better (2p).

h) The test for alpha is a test on whether the additional parameter in a negative binomial model is significantly different from 0. This indicates that the variance of the model is greater than the mean, and is used to check whether a Poisson GLM model may be sufficient or not. Here, the alpha parameter is significantly different from zero, indicating that a negative binomial model is preferred over a Poisson model (2p). The coefficient for treatment means that treatment B has 1-0.70=30% lower average length of stay (1p). The Spearman correlation indicate a very high correlation between costs and length of stay, which may not be surprising. Hence, the effects of the independent variables on length of stay also seem similar to the effects for the cost regressions (3p).

i) The marginal effects for treatment are calculated by first assuming all patients in the data are in treatment regimen A with values of other variables as observed and calculating the average predicted length of stay from the model, then by assuming all patients in the data are in treatment regimen B with values of other variables as observed and calculating the average predicted length of stay. This illustrates the independent effect of treatment reported in length of stay days, adjusted for the other variables in the model (2p). The result shows that on treatment regimen A the average predicted length of stay is 4.36 days, whereas this is 3.08 days under regimen B. This is what the relative effect of 30% reduction in length of stay in h) corresponds to in absolute terms in days (2p).

Grading: 0-19p=F

20-23p=E

24-29p=D

30-37p=C

38-44p=B

45-50p=A