

```
. ttest costs, by(treatment)
```

Two-sample t test with equal variances

Group	Obs	Mean	Std. Err.	Std. Dev.	[95% Conf. Interval]	
Treatmen	40	1639.47	265.6734	1680.266	1102.095	2176.845
Treatmen	46	1057.068	210.3731	1426.82	633.3549	1480.781
combined	86	1327.953	169.0795	1567.979	991.7775	1664.128
diff		582.4021	335.0236		-83.82903	1248.633

```
diff = mean(Treatmen) - mean(Treatmen)          t = 1.7384
Ho: diff = 0                                     degrees of freedom = 84
```

```
Ha: diff < 0                                     Ha: diff != 0                                     Ha: diff > 0
Pr(T < t) = 0.9571                             Pr(|T| > |t|) = 0.0858                             Pr(T > t) = 0.0429
```

```
.
. ranksum costs, by(treatment)
```

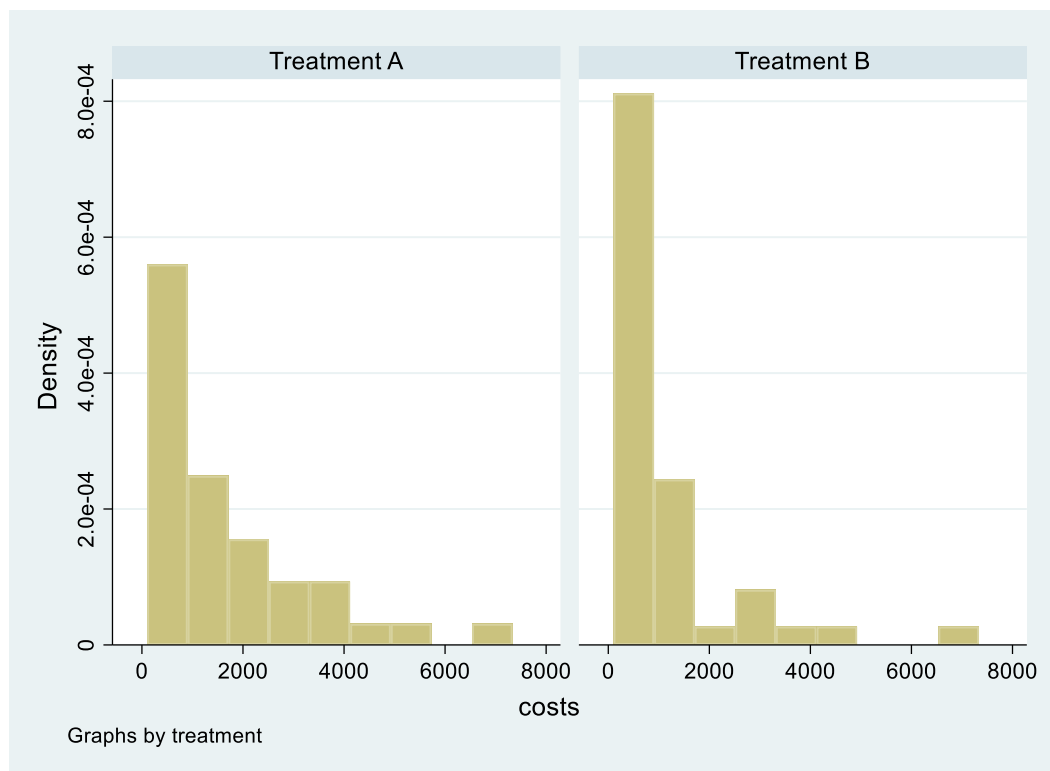
Two-sample Wilcoxon rank-sum (Mann-Whitney) test

treatment	obs	rank sum	expected
Treatment A	40	1974	1740
Treatment B	46	1767	2001
combined	86	3741	3741

```
unadjusted variance    13340.00
adjustment for ties    -0.13
```

```
adjusted variance      13339.87
```

```
Ho: costs(treatm~t==Treatment A) = costs(treatm~t==Treatment B)
      z = 2.026
      Prob > |z| = 0.0428
      Exact Prob = 0.0426
```



a) For t-test: H_0 Mean costs are equal for the treatments H_1 Mean costs are different (1p)

For Wilcoxon: H_0 Cost data come from same distributions/medians are equal H_1 Data come from different distributions/medians are different (1p)

The conclusions are conflicting: Using a 5% significance level, the t-test results in non-significant difference between the groups, while the Wilcoxon test results in a significant p-value, just below 5%. (2p) As the histograms show that the costs are very skewed (and the sample size is moderate), one would think that the Wilcoxon test produces the most reliable result. The t-test could easily be influenced by the most extreme observations. (2p)

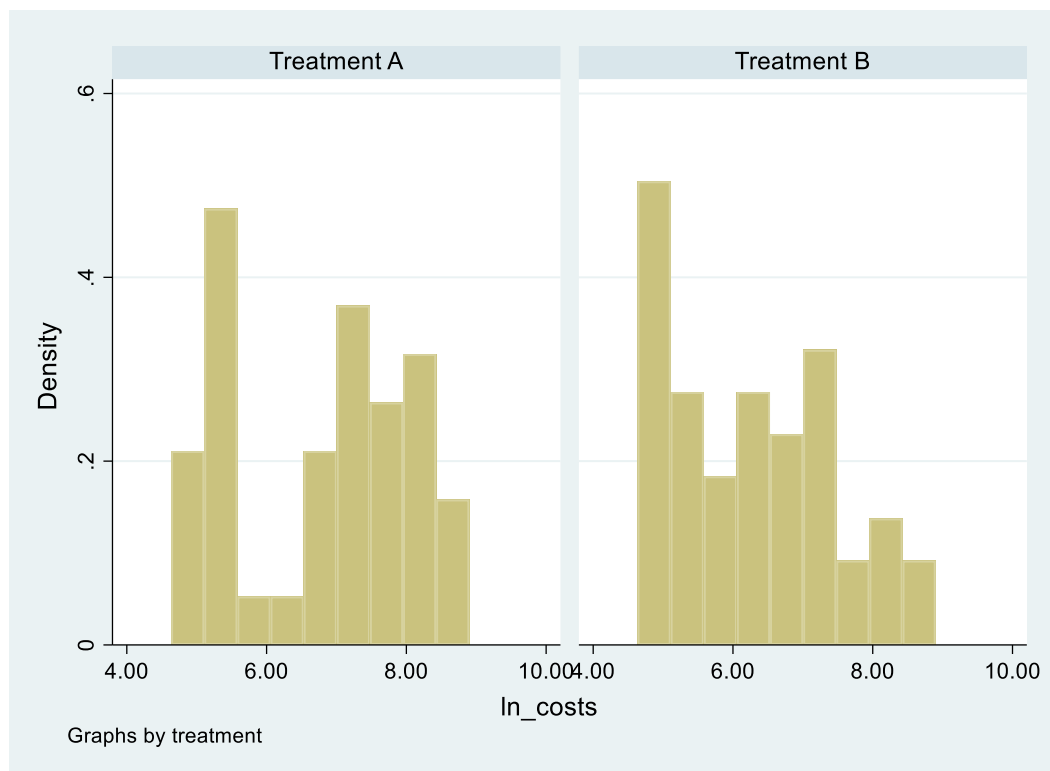
```
. ttest ln_costs, by(treatment)
```

Two-sample t test with equal variances

Group	Obs	Mean	Std. Err.	Std. Dev.	[95% Conf. Interval]	
Treatmen	40	6.781178	.1970144	1.246028	6.382678	7.179677
Treatmen	46	6.264279	.1760349	1.193927	5.909727	6.618832
combined	86	6.504697	.1335679	1.238657	6.239128	6.770266
diff		.5168984	.2634075		-.0069162	1.040713

diff = mean(Treatmen) - mean(Treatmen) t = 1.9624
 Ho: diff = 0 degrees of freedom = 84

Ha: diff < 0 Ha: diff != 0 Ha: diff > 0
 Pr(T < t) = 0.9735 Pr(|T| > |t|) = 0.0530 Pr(T > t) = 0.0265



b) We can see from the histograms that the log-transformation has eliminated the skewness. However, the conclusion for log costs is that the mean difference is just borderline significant (p-value 5.3%). (1p) The point estimate and 95% confidence interval for the difference, transformed back to the original scale, is $\exp(0.517)=1.68$ and $(\exp(-0.007), \exp(1.04))=(0.99, 2.83)$ (1p) The interpretation is that treatment B has 68% higher (geometric) mean costs than treatment A, with a confidence interval extending from just equal mean costs to 2.83 times higher costs. (2p)

. tab treatment gender, row

Key
<i>frequency</i>
<i>row percentage</i>

treatment	gender		Total
	Male	Female	
Treatment A	18 45.00	22 55.00	40 100.00
Treatment B	22 47.83	24 52.17	46 100.00
Total	40 46.51	46 53.49	86 100.00

c) (8p) Hypotheses e.g.: $H_0: p_{\text{female, treatA}} = p_{\text{female, treatB}}$ $H_1: p_{\text{female, treatA}} \neq p_{\text{female, treatB}}$ (1p)

The numbers produced by “row” tell us the percentages of male and female patients in each treatment group E.g. 45% males in treatment group A, while there are 47.8% males in treatment group B (1p). Expected frequencies:

$$40 \cdot 40 / 86 = 18.6 \quad 40 \cdot 46 / 86 = 21.4 \quad 40 \cdot 46 / 86 = 21.4 \quad 46 \cdot 46 / 86 = 24.6 \quad (1p)$$

$$\frac{(18-18.6)^2}{18.6} + \frac{(22-21.4)^2}{21.4} + \frac{(22-21.4)^2}{21.4} + \frac{(24-24.6)^2}{24.6} = 0.07 \quad (2p)$$

Should be compared to 95%-percentile of a chi-square distribution with $(2-1) \cdot (2-1) = 1$ d.f., i.e. 3.84 (1p). Accept H_0 since 0.07 is much lower than 3.84. (1p)

. spearman costs age

Number of obs = 86
Spearman's rho = 0.4259

Test of H_0 : costs and age are independent
Prob > |t| = 0.0000

d) Spearman's rho is 0.42 and statistically significant ($p < 0.01$). This tells us that there is a significant correlation between age and costs, and that costs increase with higher age (2p). The advantages compared with Pearson correlation are that, since costs are skewed distributed, Spearman's rho does not assume normality in the data. Influence of very high costs will be reduced since rankings of observations are used instead of the actual observations. Also, Spearman does not assume a strictly linear relationship between the two variables (2p).

. kwallis costs, by(severity)

Kruskal-Wallis equality-of-populations rank test

severity	Obs	Rank Sum
Low severity	38	1354.00
Medium severity	25	989.00
High severity	23	1398.00

chi-squared = 15.414 with 2 d.f.
probability = 0.0004

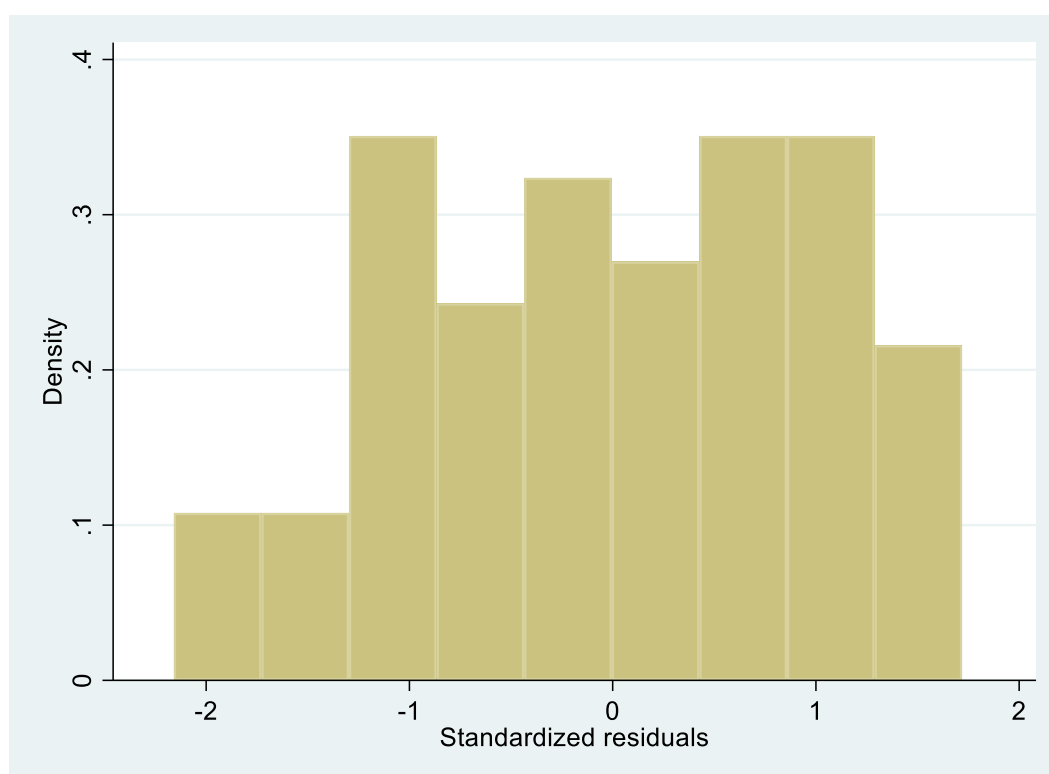
chi-squared with ties = 15.414 with 2 d.f.
probability = 0.0004

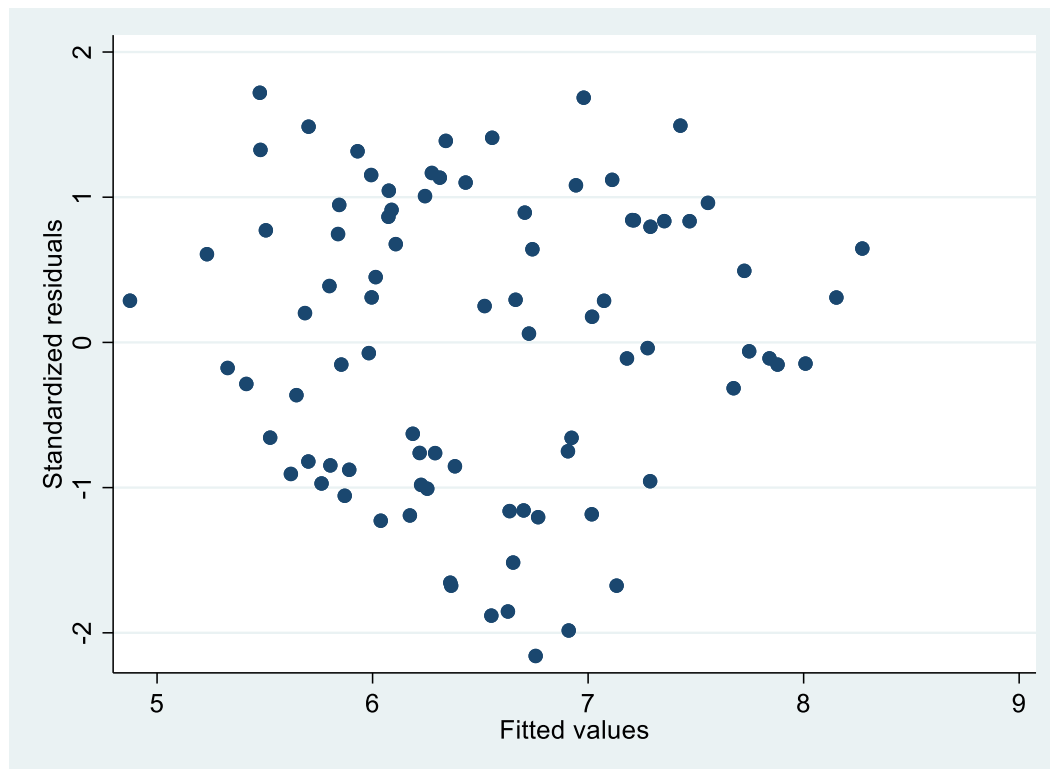
e) The null hypothesis is that cost data in all three severity groups come from the same distribution/medians are all equal, while the alternative is that at least two of the cost distributions are different/at least two medians are different (2p). Again, the test uses rankings for the observations to reduce influence of extreme observations. The test will then compare the rank sums (1p). The conclusion is to reject the null hypothesis, the p-value is < 0.01 . (1p)

```
. regress ln_costs treatment age gender i.severity, eform(coefficient)
```

Source	SS	df	MS	Number of obs	=	86
Model	47.0108188	5	9.40216376	F(5, 80)	=	9.02
Residual	83.4023279	80	1.0425291	Prob > F	=	0.0000
				R-squared	=	0.3605
				Adj R-squared	=	0.3205
Total	130.413147	85	1.53427231	Root MSE	=	1.021

ln_costs	coefficient	Std. Err.	t	P> t	[95% Conf. Interval]	
treatment	.6054614	.1337927	-2.27	0.026	.3900339	.9398761
age	1.040709	.0104432	3.98	0.000	1.020132	1.0617
gender	1.166263	.2612102	0.69	0.494	.7468332	1.82125
severity						
Medium se..	1.147229	.3065997	0.51	0.609	.6740174	1.952672
High seve..	2.956566	.811364	3.95	0.000	1.712407	5.104677
_cons	53.46128	33.34733	6.38	0.000	15.45023	184.9881





f) The purpose of the eform-option is to get coefficients reported on the anti-log scale. One will then get a relative effect interpretation of the coefficients (2p). The coefficient for treatment means that treatment B has (geometric) mean costs which are around 39.5% lower than Treatment A ($=1-0.605$) (2p). The coefficients for severity are based on using mild severity as the reference category. Patients with medium severity have 14.7% higher (geometric) mean costs compared with low severity (as 1.147 is 14.7% above 1, no difference), this is non-significant however ($p=0.61$). High severity patients have significantly ($p<0.01$) higher (geometric) mean costs, almost three times higher (2p). The effect on costs of increasing age by 15 years, will be $1.04^{15}=1.80$ or an 80% increase in the (geometric) mean costs (2p). In c) we saw that the gender distribution was almost equal across the treatment groups. This means that one expects little influence on the coefficient for treatment after adjusting for gender, regardless of whether gender has an independent effect on costs or not (2p). The histogram shows the distribution of the residuals, which should be approximately normally distributed. This looks ok, at least it is not skewed (1p). The scatter plot would indicate heteroscedasticity, or unequal variation around the fitted model as a function of high/low predicted values. This also seems fairly ok, it is a random cloud with no funnel patterns. (1p)

```
. glm costs i.treatment age gender i.severity, family(gamma) link(log) eform
```

```
Iteration 0:  log likelihood = -682.30711
Iteration 1:  log likelihood = -681.14012
Iteration 2:  log likelihood = -681.14001
Iteration 3:  log likelihood = -681.14001
```

```
Generalized linear models              Number of obs   =          86
Optimization      : ML                 Residual df     =          80
                                      Scale parameter =   .7399227
Deviance          =   71.4721163        (1/df) Deviance =   .8934015
Pearson           =   59.19381479       (1/df) Pearson  =   .7399227
```

```
Variance function: V(u) = u^2          [Gamma]
Link function      : g(u) = ln(u)       [Log]
```

```
Log likelihood    = -681.1400104        AIC           =          15.98
                                      BIC           =   -284.8757
```

costs	OIM					
	exp(b)	Std. Err.	z	P> z	[95% Conf. Interval]	
treatment						
Treatment B	.6013947	.1119144	-2.73	0.006	.4175979	.8660859
age	1.042291	.0095841	4.50	0.000	1.023674	1.061245
gender	1.107742	.213189	0.53	0.595	.7596641	1.615308
severity						
Medium se..	1.027929	.2366042	0.12	0.905	.6546923	1.613946
High seve..	2.571347	.6315265	3.85	0.000	1.588928	4.161184
_cons	81.79454	45.53803	7.91	0.000	27.46813	243.5676

```
. margins treatment
```

```
Predictive margins                                Number of obs    =          86  
Model VCE      : OIM
```

```
Expression    : Predicted mean costs, predict()
```

	Delta-method					
	Margin	Std. Err.	z	P> z	[95% Conf. Interval]	
treatment						
Treatment A	1685.455	254.5174	6.62	0.000	1186.611	2184.3
Treatment B	1013.624	145.4956	6.97	0.000	728.4578	1298.79

```
.  
. margins, dydx(treatment)
```

```
Average marginal effects                        Number of obs    =          86  
Model VCE      : OIM
```

```
Expression    : Predicted mean costs, predict()  
dy/dx w.r.t.  : 1.treatment
```

	Delta-method					
	dy/dx	Std. Err.	z	P> z	[95% Conf. Interval]	
treatment						
Treatment B	-671.8314	266.3701	-2.52	0.012	-1193.907	-149.7556

Note: dy/dx for factor levels is the discrete change from the base level.

```
. margins severity
```

```
Predictive margins                                Number of obs    =          86
Model VCE      : OIM
```

```
Expression    : Predicted mean costs, predict()
```

	Delta-method					
	Margin	Std. Err.	z	P> z	[95% Conf. Interval]	
severity						
Low sever..	900.8031	151.3563	5.95	0.000	604.1501	1197.456
Medium se..	925.9615	164.0539	5.64	0.000	604.4218	1247.501
High seve..	2316.277	422.6333	5.48	0.000	1487.931	3144.623

```
.
. margins, dydx(severity)
```

```
Average marginal effects                        Number of obs    =          86
Model VCE      : OIM
```

```
Expression    : Predicted mean costs, predict()
```

```
dy/dx w.r.t.  : 1.severity 2.severity
```

	Delta-method					
	dy/dx	Std. Err.	z	P> z	[95% Conf. Interval]	
severity						
Medium se..	25.15843	210.4357	0.12	0.905	-387.288	437.6049
High seve..	1415.474	446.1106	3.17	0.002	541.1131	2289.834

Note: dy/dx for factor levels is the discrete change from the base level.

g) The magnitude of the coefficient is similar to the log-linear model, yielding an estimate of around 40% lower mean costs treatment B (1p). However, the GLM model is comparing the standard arithmetic mean costs, not geometric mean costs as in the log-linear model (2p). The means produced by the margins command are not equal to those produced by the t-test, since margins show the adjusted predicted means. That is, comparing patients of the same age, gender and severity (2p). The margins for severity shows first what the predicted mean costs are for each severity group. Again, these are adjusted for the other variables in the model. The second margins translate the relative adjusted effects in the regression table to cost differences on the standard scale. We see for instance that patients with high severity have around 1415 kroner higher costs than those with low severity (2p). The p-values are different simply because the dydx-table shows p-values on whether the medium/high groups are significantly different from the low severity group, while the other margins table show p-values on whether the predicted mean costs are significantly different from 0 (2p).

```
. lgamma costs treatment gender age medsev highsev, eform
```

Fitting constant-only model:

```
Iteration 0: log likelihood = -114203.94
Iteration 1: log likelihood = -1005.767
Iteration 2: log likelihood = -714.90403
Iteration 3: log likelihood = -707.63737
Iteration 4: log likelihood = -703.76218
Iteration 5: log likelihood = -703.74707
Iteration 6: log likelihood = -703.74707
```

Fitting full model:

```
Iteration 0: log likelihood = -703.74707
Iteration 1: log likelihood = -688.20052
Iteration 2: log likelihood = -679.05203
Iteration 3: log likelihood = -678.99452
Iteration 4: log likelihood = -678.99448
```

```
Log-gamma model          Number of obs   =          86
                          LR chi2(5)        =          49.51
Log likelihood = -678.99448  Prob > chi2      =          0.0000
```

costs	IRR	Std. Err.	z	P> z	[95% Conf. Interval]	
treatment	.6013944	.1121611	-2.73	0.006	.417262	.8667821
gender	1.107743	.2136592	0.53	0.596	.7590333	1.616653
age	1.042291	.0096052	4.49	0.000	1.023634	1.061287
medsev	1.027929	.2371259	0.12	0.905	.6540417	1.615552
highsev	2.571346	.6329188	3.84	0.000	1.587243	4.165602
_cons	81.79442	45.63836	7.89	0.000	27.40209	244.1539

Note: Estimates are transformed only in the first equation.

/ln_phi	-.2968026	.1376706	-2.16	0.031	-.566632	-.0269732

phi	.7431907	.1023155			.5674333	.9733874

```
. estat ic
```

Akaike's information criterion and Bayesian information criterion

Model	N	ll(null)	ll(model)	df	AIC	BIC
.	86	-703.7471	-678.9945	7	1371.989	1389.169

Note: BIC uses N = number of observations. See [\[R\] BIC note](#).

```
. regress ln_costs treatment age gender i.severity, eform(coefficient)
```

Source	SS	df	MS	Number of obs	=	86
Model	47.0108188	5	9.40216376	F(5, 80)	=	9.02
Residual	83.4023279	80	1.0425291	Prob > F	=	0.0000
				R-squared	=	0.3605
				Adj R-squared	=	0.3205
Total	130.413147	85	1.53427231	Root MSE	=	1.021

ln_costs	coefficient	Std. Err.	t	P> t	[95% Conf. Interval]	
treatment	.6054614	.1337927	-2.27	0.026	.3900339	.9398761
age	1.040709	.0104432	3.98	0.000	1.020132	1.0617
gender	1.166263	.2612102	0.69	0.494	.7468332	1.82125
severity						
Medium se..	1.147229	.3065997	0.51	0.609	.6740174	1.952672
High seve..	2.956566	.811364	3.95	0.000	1.712407	5.104677
_cons	53.46128	33.34733	6.38	0.000	15.45023	184.9881

```
.
. summarize(ln_costs)
```

Variable	Obs	Mean	Std. Dev.	Min	Max
ln_costs	86	6.504697	1.238657	4.634729	8.900731

```
.
. display 2*(-(e(ll)-r(sum))+7)
1374.2276
```

h) The output shows AIC values of 1371 for the gamma GLM model, while the AIC for the log-linear model is 1374. As a lower value indicate a better fit, the gamma GLM model performs slightly better (2p). However, even though we have the interpretational difference between geometric and arithmetic means etc., we see that all coefficients are quite similar. They all have the same sign, the conclusions as to which variables are significant are the same, and the plots in f) indicated no great issues with the log-linear model either. Hence, the choice of model seems not to matter much in practice (2p).