Homework 2

a) $\bar{x} = 26.676$

$s = 12.839$

The population is assumed to be normal and the true mean and variance are unknown, thus I use s as an estimator for σ.

confidence interval $= (\bar{x} - c\frac{s}{\sqrt{n}}, \bar{x} + c\frac{s}{\sqrt{n}})$ with $P(T_{n-1} \geq c) = \frac{1}{2}\alpha$

$\frac{\alpha}{2} = \frac{0.05}{2} = 0.025$

degrees of freedom $= n - 1 = 24 - 1 = 23$

Now, I use the table for the t-distribution to find $c = 2.069$

confidence interval $= (26.676 - 2.069 \frac{12.839}{\sqrt{24}}, 26.676 + 2.069 \cdot \frac{12.839}{\sqrt{24}}) =$

$= (21.254, 32.098)$

b) The interpretation of the interval in a. (in words) is: "We are 95% confident that the expected mean is between 21.254 and 32.098". Because this interval contains 30, I cannot decide whether μ is bigger, smaller or equal to 30. The only way I could conclude that the claim was wrong would be if the entire interval would be above 30, which is not the case. So, I believe the claim is reasonable.

c) The histogram shows a shape that is very consistent with the exponential distribution and discards the possibility of a normal distribution. The same can be argued for the box-plot, the shape is consistent with the exponential distribution. The normal QQ-plot shows a strong deviation from the line $y=x$, the exponential QQ-plot seems like a better fit. The visual analysis based on the histogram, box-plot and QQ-plot suggest this data comes from an exponential distribution, thus I consider that the normality assumption is NOT justified.

d) The histogram shows the typical bell-shape that is associated with the normal distribution. The shape of the box-plot is more consistent with the normal distribution than the exponential one. The QQ-plot shows very little to NO deviation from the line $y=x$, the normal QQ-plot seems like a good fit. The visual analysis based on the histogram, box-plot and QQ-plot suggest this data comes from a normal distribution, thus I consider that the normality assumption is justified.

e)

f) The interpretation of the interval in e.(in words) is: "We are 95% confident that the expected median is between 19.906 and 29.224". Because this interval doesn't contain 30, I can assume that the median is smaller than 30. So, I believe the claim that "the median of the task times is smaller than 30 minutes" is justified.

g) To construct a 90% confidence interval we need to find a new value of c which satisfies $P(T_{49} \geq c) = 0.05$. From the table, we get: $c = 1.676$. The new confidence interval is: $(26.676 - 1.676 \frac{12.839}{\sqrt{50}}, 26.676 + 1.676 \frac{12.839}{\sqrt{50}}) =$

$= (23.633, 29.719)$

h) We are 90% confident that the expected median is between 23.633 and 29.719.

90% confident means that out of 100 tries/computed values, 90 values are in the stated interval.