

Task p. 27-2 / Anwendung S. 27-2

FK

automatic

Working directory

```
> setwd("D:/kronthafranz/Documents/01Lehre/06Quantitative Forschungsmethoden  
dt en")
```

Load data

```
> load("D:/kronthafranz/Documents/01Lehre/06Quantitative Forschungsmethoden  
dt en/04Parametric and nonparametric testing/gssft.RData")
```

Define factor

```
> gssft <- within(gssft, {  
+   f_sex <- as.factor(sex)  
+ })
```

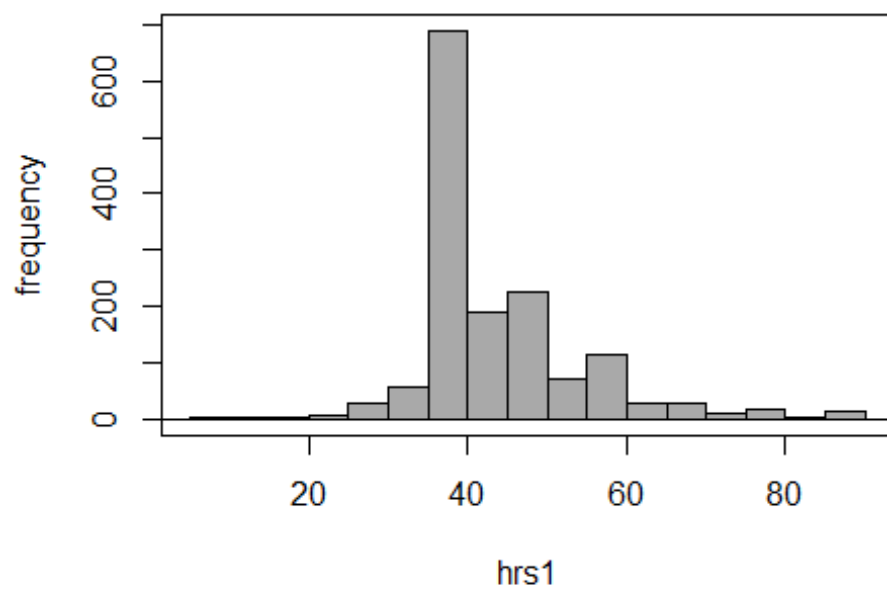
Descriptive statistics

```
> numSummary(gssft[, "hrs1", drop=FALSE], groups=gssft$f_sex,  
statistics=c("mean", "sd",  
+   "quantiles", "skewness", "kurtosis"), quantiles=c(0,.25,.5,.75,1),  
type="2")
```

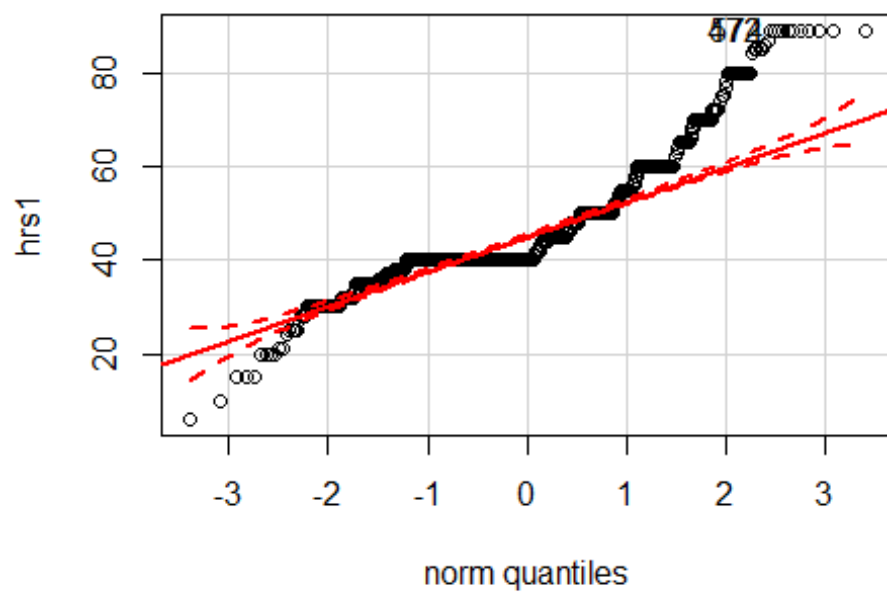
	mean	sd	skewness	kurtosis	0%	25%	50%	75%	100%	hrs1:n	hrs1:NA
1	47.83333	11.22970	1.204486	2.173362	6	40	45	52	89	792	2
2	43.14894	9.321999	1.698619	5.699021	10	40	40	45	89	705	1

Normal distribution

```
> with(gssft, Hist(hrs1, scale="frequency", breaks="Sturges",  
col="darkgray"))
```



```
> with(gssft, qqPlot(hrs1, dist="norm", id.method="y", id.n=2,
labels=rownames(gssft)))
```



```
472 574
1487 1488
```

```
> normalityTest(~hrs1, test="shapiro.test", data=gssft)
```

Shapiro-Wilk normality test

data: hrs1

W = 0.83951, p-value < 2.2e-16

Same variance

```
> with(gssft, tapply(hrs1, f_sex, var, na.rm=TRUE))
```

```
      1      2
126.10619  86.89966
```

```
> leveneTest(hrs1 ~ f_sex, data=gssft, center="mean")
```

Levene's Test for Homogeneity of Variance (center = "mean")

	Df	F value	Pr(>F)
group	1	36.748	0.000000001699 ***
	1495		

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

t-Test independent sample (unequal variance)

```
> t.test(hrs1~f_sex, alternative='two.sided', conf.level=.95,
var.equal=FALSE, data=gssft)
```

Welch Two Sample t-test

data: hrs1 by f_sex

t = 8.8136, df = 1487.9, p-value < 2.2e-16

alternative hypothesis: true difference in means is not equal to 0

95 percent confidence interval:

3.641838 5.726956

sample estimates:

mean in group 1	mean in group 2
47.83333	43.14894

Wilcoxon Rank-Sum Test

```
> with(gssft, tapply(hrs1, f_sex, median, na.rm=TRUE))
```

```
 1  2
45 40
```

```
> wilcox.test(hrs1 ~ f_sex, alternative="two.sided", data=gssft)
```

Wilcoxon rank sum test with continuity correction

data: hrs1 by f_sex

W = 361500, p-value < 2.2e-16

alternative hypothesis: true location shift is not equal to 0

Interpretation:

Variable is not normally distributed -> nonparametric test

Large sample -> parametric test is relatively robust with regard to the violation of the normal distribution assumption

Both tests lead to valid results

H0 is rejected, there is a difference in working time between men and women