Answer 12.44

Use LEAD.DAT data

Use nonparametric methods to compare MAXFWT among the three exposure groups defined by the variable LEAD_GRP.

Additionally, Assess the distribution of Maxfwt to verify that non-parametric methods are needed. (Discuss why or why not) regardless of your answer, use non-parametric methods. In practice, we would only perform testing for pairwise differences if the overall test is significant.

We first assess the normality with Shapiro-Wilk test by group.

```
. by lead_grp, sort : swilk maxfwt
```

-> lead_grp = 1

Shapiro–Wilk W test for normal data

Variable	Obs	W	V	z	Prob>z
maxfwt	78	0.88747	7.565	4.427	0.00000

-> lead_grp = 2

Shapiro–Wilk W test for normal data

Variable	Obs	W	V	z	Prob>z
maxfwt	24	0.86970	3.515	2.563	0.00519

-> lead_grp = 3

Shapiro–Wilk W test for normal data

Variable	Obs	W	V	z	Prob>z
maxfwt	22	0.94754	1.329	0.577	0.28204

The results showed that Group 1 is not normally distributed. Therefore, we should use Kruskal–Wallis rather than ANOVA.

We then check for ties.

```
. duplicates report maxfwt

Duplicates in terms of maxfwt

    Copies │ Observations        Surplus
   ────────┼──────────────────────────────
       1   │        16               0
       2   │        18               9
       3   │        18              12
       4   │        16              12
       5   │         5               4
       6   │        18              15
       8   │         8               7
      25   │        25              24
```

The results show that we have ties in the data. We will use Kruskal–Wallis with ties.

```
. kwallis maxfwt, by(lead_grp)

Kruskal–Wallis equality-of-populations rank test

   ┌─────────┬──────┬──────────┐
   │ lead_grp│ Obs  │ Rank sum │
   ├─────────┼──────┼──────────┤
   │    1    │  78  │ 5174.00  │
   │    2    │  24  │ 1178.00  │
   │    3    │  22  │ 1398.00  │
   └─────────┴──────┴──────────┘

   chi2(2) =   4.251
      Prob = 0.1194

   chi2(2) with ties =   4.290
               Prob = 0.1171
```

The p-value is 0.1171 > 0.05, thus the test is not significant. We can conclude that we do not find significance in MAXFWT across three exposure groups.

Required additional problems
Using LEAD.DAT.dta

a) Test whether there are any differences in full IQ score across the 3 exposure groups defined by lead_grp (no exposure, past exposure and current exposure). For homework purposes, do so both parametrically (hint: refer to ANOVA from 1690 course to perform the parametric test or Rosner chapter 12; you can also refer to the ANOVA section from the STATA lab) and non-parametrically. State the hypothesis, report the test statistics and p value

The hypothesis is: H_0: *the mean of full IQ score is equal across* 3 *groups*, H_1: *At least one mean of full IQ score differs*.

We first use ANOVA test.

```
. oneway iqf lead_grp
```

```
                        Analysis of variance
     Source               SS         df      MS              F     Prob > F

Between groups        710.868374      2    355.434187       1.73     0.1810
 Within groups        24808.3252    121    205.027481

     Total            25519.1935    123    207.473118

Bartlett's equal-variances test: chi2(2) =    4.9875     Prob>chi2 = 0.083
```

The p-value is 0.083 > 0.05, thus the test is not significant. We cannot reject the null hypothesis.

We then use the nonparametric method.
```
. duplicates report iqf

Duplicates in terms of iqf
```

Copies	Observations	Surplus
1	20	0
2	22	11
3	12	8
4	32	24
5	5	4
6	18	15
7	7	6
8	8	7

```
. kwallis iqf, by(lead_grp)

Kruskal-Wallis equality-of-populations rank test
```

lead_grp	Obs	Rank sum
1	78	5228.50
2	24	1328.00
3	22	1193.50

```
  chi2(2) =  3.354
    Prob = 0.1869

  chi2(2) with ties =  3.358
             Prob = 0.1866
```

The p-value is 0.1866>0.05. Therefore, the test is not significant and we conclude that the differences between the means are not significant.

b) Assess the distribution of full IQ scores and discuss which test is more valid and justify why.
```
. by lead_grp, sort : swilk iqf
```

```
-> lead_grp = 1
```

Shapiro-Wilk W test for normal data

Variable	Obs	W	V	z	Prob>z
iqf	78	0.98241	1.182	0.367	0.35690

```
-> lead_grp = 2
```

Shapiro-Wilk W test for normal data

Variable	Obs	W	V	z	Prob>z
iqf	24	0.95111	1.319	0.564	0.28629

```
-> lead_grp = 3
```

Shapiro-Wilk W test for normal data

Variable	Obs	W	V	z	Prob>z
iqf	22	0.93904	1.544	0.881	0.18910

All three groups are normally distributed. Therefore, the parametric test is more valid.

c) Given your results to the overall test, is there sufficient evidence to perform further pairwise testing between the means (yes or no, and explain why or why not, in 1 or 2 sentences.) For each test performed, state the null and alternative hypotheses, report the test statistic, and p-value, and interpret your tests.

Since the result of Bartlett's test is not significant, there is no need to perform pairwise test between the means.