MATH60604A *Statistical Modelling*

Midterm examination

Exam booklet

Practice exam

Léo Belzile

Instructions: The time allotted for the examination is 180 minutes. A non-programmable calculator may be used. You are allowed a single-sided crib sheet.

There are a total of 40 marks available in the exam paper, the distribution of which can be found in the right margin.

You must hand back the **exam booklet** at the end of the examination.

Last name:

First name:

STUDENT ID:

Question:	1	2	3	Total
Points:	12	12	16	40
Score:				

Question 1. .... $\boxed{12}$

In extreme value analysis, large-sample theory dictates that exceedances of Y above a high threshold u are well approximated by a **generalized Pareto** distribution $Z = Y - u \stackrel{\cdot}{\sim} \mathsf{GP}(\sigma, \xi)$ with scale $\sigma > 0$ and shape $\xi \in \mathbb{R}$, with distribution and density functions

$$F(z) = 1 - (1 + \xi z/\sigma)_{+}^{-1/\xi}, \qquad f(z) = \sigma^{-1}(1 + \xi z/\sigma)_{+}^{-1/\xi - 1},$$

with $(x)_+ = \max\{x, 0\}$; the case $\xi = 0$ is defined by continuity (exponential submodel).

We consider the largest fire insurance claims (in millions Danish krones) filed at the Copenhagen Re company in Denmark from January 1980 until the end of December 1990 ($n_Y = 11$ years worth of data). We model the $n = 109$ exceedances above 10 millions krones, corresponding to $\zeta = 0.0503$ proportion of the data. Our objective is to provide 100-years return levels for risk analysis.

The T-year return level r_T, is a high quantile exceeded with tail probability p, where $p = \zeta n_Y / T$, with ζ the proportion of observations above the threshold, n_Y the number of years worth of data and $T = 100$ the number of year. If we invert the distribution function, we find

$$r_T = \frac{\sigma}{\xi}\left\{(\zeta n_Y / T)^{-\xi} - 1\right\}$$

1.1 Write the log likelihood function for an n-sample of independent threshold exceedances z_i, $(i = 1, \ldots, n)$ for $\xi \neq 0$. [2]

Solution: Assuming $\xi \neq 0$, the joint density is

$$L(\sigma, \xi) = f(\boldsymbol{z} = \prod_{i=1}^{n} f(z_i) = \sigma^{-n} \prod_{i=1}^{n} (1 + \xi z_i/\sigma)^{-1/\xi - 1}$$

and so the log likelihood

$$\ell(\sigma, \xi) = -n \ln(\sigma) - \left(1 + \frac{1}{\xi}\right) \sum_{i=1}^{n} \ln(1 + \xi z_i/\sigma)$$

1.2 If we reparametrize the model in terms of ξ and $\theta = \xi/\sigma$, show that we can derive an explicit formula for the profile log likelihood $\ell_{\mathrm{p}}(\theta)$, thereby reducing the optimization to a one-dimensional problem. [2]

Solution: With the substitution,

$$\ell(\theta,\xi) = n\ln(\theta) - n\ln(\xi) - \left(1+\frac{1}{\xi}\right)\sum_{i=1}^{n}\ln(1+\theta z_i)$$

and for fixed θ_0,

$$\left.\frac{\partial \ell(\theta,\xi)}{\partial \xi}\right|_{\theta=\theta_0} = -\frac{n}{\xi} + \frac{1}{\xi^2}\sum_{i=1}^{n}\ln(1+\theta_0 z_i) = 0$$

so $\widehat{\xi}_\theta = n^{-1}\sum_{i=1}^{n}\ln(1+\theta z_i)$. The second derivative is

$$\left.\frac{\partial^2 \ell(\theta,\xi)}{\partial \xi^2}\right|_{\theta=\theta_0} = \frac{n}{\xi^2} - \frac{2}{\xi^3}\sum_{i=1}^{n}\ln(1+\theta_0 z_i)$$

and evaluating the latter at $\widehat{\xi}_\theta$ gives $-n/\widehat{\xi}_\theta^2 < 0$, so this is indeed the maximum.
The profile log likelihood is

$$\ell_{\mathrm{p}}(\theta) = n\ln\theta - n\ln\widehat{\xi}_\theta - \left(1+\frac{1}{\widehat{\xi}_\theta}\right)\sum_{i=1}^{n}\ln(1+\theta z_i).$$

```
Log-likelihood: -374.893
Sample size: 109
Proportion above threshold: 0.0503

Estimates                    Standard errors
scale   shape                scale    shape
6.975   0.497                1.1135   0.1363
```

Listing 1: Maximum likelihood estimate for the generalized Pareto

1.3 An optimization routine has yielded the following estimates for the MLE in Listing 1. Explain how you could test whether $\xi = 0$ (exponential model) using [2]

(a) a Wald test using the output from Listing 1 and

(b) a likelihood ratio test if in addition

```
> sum(dexp(y, rate = 1/mean(y), log = TRUE))
-397.2921
```

Listing 2: Code for an exponential log likelihood

Solution:

- Fit the exponential submodel (whose MLE is the sample mean) to get log likelihood value $\ell_0(\widehat{\sigma}_0, 0)$. Compute the likelihood ratio statistic $R = 2\{\ell(\widehat{\sigma}, \widehat{\xi}) - \ell_0(\widehat{\sigma}_0, 0)\}$, with $\ell(\widehat{\sigma}, \widehat{\xi})$ given in the output and compare the numerical value to a χ^2_1 distribution.
- Compute the Wald statistic $W = (\widehat{\xi} - 0)/\mathrm{se}(\widehat{\xi}) = 0.497/0.1363 = 3.646$ and compare to a standard normal (reject $\mathscr{H}_0 : \xi = 0$ at level 5%, the data are not from the exponential submodel.

1.4 One can show that the Fisher information matrix is [2]

$$\imath(\sigma,\xi) = n\begin{pmatrix} \sigma^{-2}(1+2\xi)^{-1} & \sigma^{-1}(1+\xi)^{-1}(1+2\xi)^{-1} \\ \sigma^{-1}(1+\xi)^{-1}(1+2\xi)^{-1} & 2(1+\xi)^{-1}(1+2\xi)^{-1}\end{pmatrix}$$

Explain how you could use the above result to derive standard errors for the parameters σ and ξ.

Solution: Plug-in the values of the MLE in the information matrix and compute the square root of the diagonal elements of $\imath^{-1}(\widehat{\sigma},\widehat{\xi})$ to get the standard errors.

1.5 Give the maximum likelihood estimate of the 100 years return level r_{100} for the Danish insurance data. [2]

Solution: The MLE is invariant, so plug-in the values in the formula for r_{100} with $T = 100$, $n_Y = 11$, $\zeta = 0.0503$ and $\widehat{\sigma} = 6.975$, $\widehat{\xi} = 0.497$ to get $\widehat{r}(100) = 171.72$ millions Danish krones.

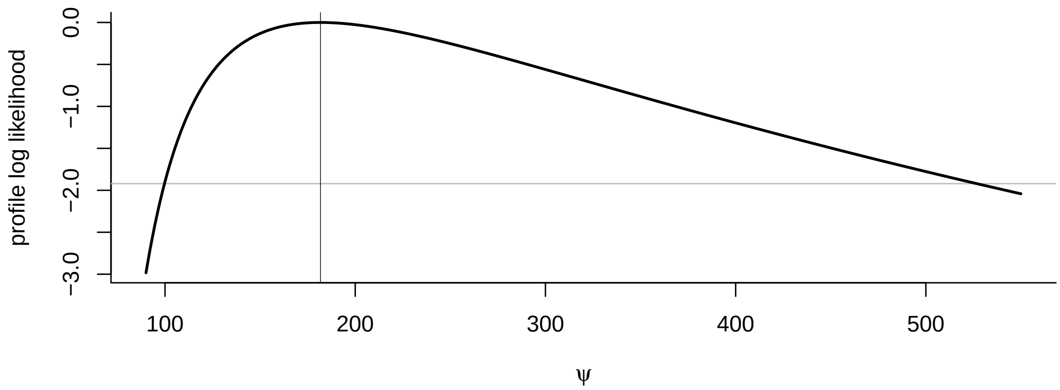

Figure 1: Profile log likelihood for the return level. The profile has been shifted to equal zero at the MLE, and the gray horizontal line indicates the cutoff for 95% profile confidence intervals based on the asymptotic χ^2_1 distribution.

1.6 Figure 1 shows the profile log likelihood for the 100-year return level $\psi = r_{100}$, with the gray horizontal line indicating the cutoff for 95% profile confidence intervals based on the asymptotic χ^2_1 distribution. Would a Wald-based confidence intervals be similar looking? [2]

Solution: The sampling distribution illustrated by the profile likelihood is quite skewed, so a symmetric 95% Wald confidence interval centered around the MLE would underestimate drastically the risk. The profile likelihood shows an interval of roughly $(100, 550)$, very different from what the Wald interval would look like.

Question 2. 12

Grossmann and Kross (2014) study the question "Are people wiser when reflecting on other people's problems compared with their own?". They "randomly assigned participants to

1. reason about their own problem from an immersed perspective (self-immersed condition),
2. reason about their friend's problem from an immersed perspective (other-immersed condition),
3. reason about their own problem from a distanced perspective (self-distanced condition), or
4. reason about their friend's problem from a distanced perspective (other-distanced condition)".

The study below also considered `age` as a separate factor.

The variables include

- `limits`: response variable, the mean-centered score for the question on the "recognition of limits of knowledge".
- `target`: factor, is the target `self` or `other`.
- `perspective`: factor, either `immersed` or `distanced`.
- `age`: age group, either `young` (20–40 years old) or `old`, (60–80 years old).

Table 1: Coefficients and standard errors for the three-way full factorial model.

	coef.	std. error
(Intercept)	0.207	0.144
target [self]	0.024	0.204
perspective [immersed]	−0.121	0.203
age [old]	−0.225	0.216
target [self]:perspective [immersed]	−0.211	0.284
target [self]:age [old]	0.059	0.308
perspective [immersed]:age [old]	0.020	0.306
target [self]:perspective [immersed]:age [old]	−0.195	0.433

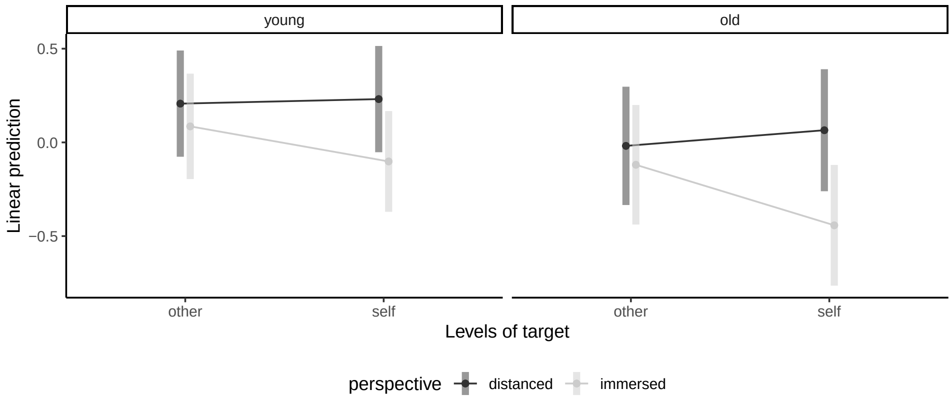

Figure 2: Estimated marginal means with 95% confidence intervals for each subgroup

Table 2: Analysis of variance table (type 2 decomposition) for the full factorial three-way ANOVA model.

	sum of squares	df	F-stat	p-value
target	1.12	1	0.87	0.35
perspective	7.59	1	5.88	0.02
age	6.12	1	4.74	0.03
target:perspective	2.45	1	1.90	0.17
target:age	0.04	1	0.03	0.86
perspective:age	0.16	1	0.13	0.72
target:perspective:age	0.26	1	0.20	0.65
residuals	570.17	442		

Table 3: Estimated marginal means for the four experimental conditions.

perspective	target	emmean	std. error	lower CL	upper CL
distanced	other	0.09	0.11	−0.12	0.31
immersed	other	−0.02	0.11	−0.23	0.20
distanced	self	0.15	0.11	−0.07	0.36
immersed	self	−0.27	0.11	−0.48	−0.06

Table 4: Marginal contrasts for the two-way marginal model.

contrast	estimate	std. error	t-stat	p-value
C_1	0.311	0.131	2.366	0.018
C_2	−0.109	0.134	−0.818	0.414
C_3	0.420	0.153	2.742	0.006
C_4	0.111	0.153	0.727	0.468

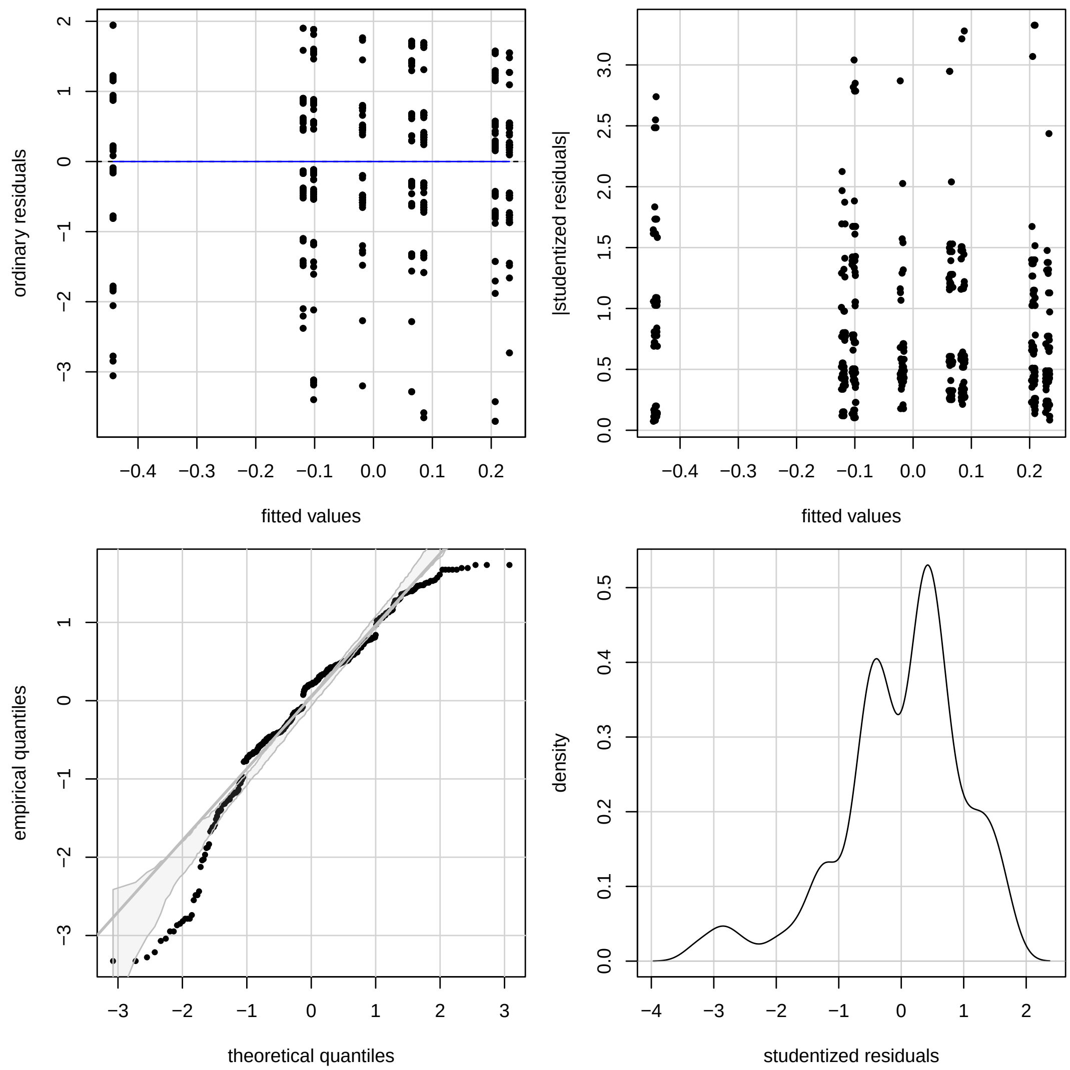

Figure 3: Diagnostic plots for the three-way ANOVA model: scatterplots of fitted versus ordinary residuals (top left), fitted vs absolute value of externally studentized residuals (top right), Student quantile-quantile plot of externally studentized residuals (bottom left) and density plot of the externally studentized residuals (bottom right).

2.1 Interpret the intercept parameter of the full factorial model fitted in Table 1. [2]

> **Solution:** The average limits score for the reference category (young, distanced, other)

2.2 Based on Table 2, how many participants were recruited in the study? [2]

> **Solution:** There are eight subgroups (coefficients) in Table 1, and 442 degrees of freedom for the residuals (Table 2), so $n = 450$ participants overall.

2.3 Based on the analysis of variance reported in Table 2, is it correct to marginalize over age and consider the main effects of (target, perspective) by pooling data for both young and old? What is the benefit of doing so? [2]

> **Solution:** Yes, because neither the three-way nor the two-way interactions involving "age" are statistically significant, so we can look at the marginal effects. Pooling data increases the sample size, thereby increasing power to detect difference between factor levels.

2.4 Write down the weight vectors for the following four contrasts, given the order of the categories in Table 3: [2]

C_1 (______, ______, ______, ______) other (immersed and distance) vs self-immersed
C_2 (______, ______, ______, ______) other (immersed and distanced) vs self-distanced
C_3 (______, ______, ______, ______) self-distanced vs self-immersed
C_4 (______, ______, ______, ______) other-distanced vs other-immersed

> **Solution:** Constrast weights must proportional to $C_1 = (1,1,0,-2)$, $C_2 = (1,1-2,0)$, $C_3 = (0,0,1,-1)$ and $C_4 = (1,-1,0,0)$.

2.5 Based on the output of Table 4, comment on the contrast analysis. [2]

Solution: Contrasts C_1 and C_3, which involve self-immersed vs rest, show that thinking about others (or asking people to self-distance) leads to significant differences, in line with the authors claims.

2.6 Based on Figure 3, what model assumption is violated? Justify your answer and discuss the impacts on inference. [2]

- independence
- incorrect mean model specification
- additivity
- homoscedasticity
- normality
- absence of outliers

Solution: The normality assumption is violated; we can see clearly the presence of a mixture distribution, with multiple modes (bottom right). This won't affect the results of tests much (although the sample isn't overly large), but the prediction intervals may not have good coverage.

Question 3. 16

Think-aloud and the impact of learning to read. The data from Baumann *et al.* (1992) study the effect of different instruction methods for reading comprehension. The data we consider are the following

- `pretest2`: score (out of 15) on second pretest using a comprehension monitoring questionnaire
- `posttest2`: response, score (out of 18) on an expanded comprehension monitoring questionnaire
- `group`: experimental group, one of directed reading-thinking activity (DRTA), think-aloud (TA) and directed reading group (DR)

We fit the models for the `posttest2` score with the **sum-to-zero** constraint for the categorical variable experimental `group`. The model matrix for the dummies is given in Table 5

	(Intercept)	group1	group2
DR	1	1	0
DRTA	1	0	1
TA	1	−1	−1

Table 5: Model matrix with the sum-to-zero constraint.

```
> model1 <- lm(posttest2 ~ group, data = BSJ92)
> model2 <- lm(posttest2 ~ offset(pretest2) + group, data = BSJ92)
> model3 <- lm(posttest2 ~ pretest2 + group, data = BSJ92)
> model4 <- lm(posttest2 ~ pretest2 * group, data = BSJ92)
```

Listing 3: R syntax for the four fitted models

```
> anova(model3, model4)
Analysis of Variance Table

Model 3: posttest2 ~ pretest2 + group
Model 4: posttest2 ~ pretest2 * group
  Res.Df    RSS Df Sum of Sq      F Pr(>F)
1     62 332.19
2     60 331.07  2    1.1264 0.1021 0.9031
```

Listing 4: Comparison of two models

Table 6: Coefficients and standard errors for different linear regression models.

	estimate	std. error
(Intercept)	6.712	0.293
group1	−1.167	0.414
group2	−0.485	0.414

(a) Coefficients for Model 1 for the post-test score 2 as a function of the experimental group (with sum-to-zero parametrization).

	estimate	std. error
(Intercept)	5.301	0.722
pretest2	0.276	0.130
group1	−1.213	0.404
group2	−0.481	0.403

(b) Coefficients for Model 2 for the post-test score 2 as a function of the experimental group (with sum-to-zero parametrization) and `pretest2`.

	estimate	std. error
(Intercept)	5.398	0.765
pretest2	0.256	0.140
group1	−1.619	0.994
group2	−0.236	1.113
pretest2:group1	0.079	0.176
pretest2:group2	−0.047	0.204

(c) Coefficients for Model 3 for the post-test score 2 as a function of the experimental group (with sum-to-zero parametrization), `pretest2` and their interaction.

Table 7: Estimated marginal means for Model 3.

group	emmean	std. error	df	lower CL	upper CL
DR	5.499	0.494	62	4.512	6.487
DRTA	6.231	0.494	62	5.245	7.218
TA	8.406	0.494	62	7.418	9.393

Table 8: Pairwise contrasts based on marginal means of Model 3.

contrast	estimate	std. error	df	t-stat.	p-value
DR − DRTA	−0.732	0.698	62	−1.048	0.550
DR − TA	−2.906	0.699	62	−4.157	$< 10^{-3}$
DRTA − TA	−2.174	0.698	62	−3.114	0.008

3.1 Compute the sample mean of each group based on Table 6. [2]

Solution: We look at Table 6 (a): the mean of group DR is $\widehat{\mu}_{\text{DR}} = 6.712 - 1.167 = 5.54$, that of DRTA is $\widehat{\mu}_{\text{DRTA}} = 6.712 - 0.485 = 6.227$ and finally for TA $\widehat{\mu}_{\text{TA}} = 6.712 + 1.167 + 0.485 = 8.364$.

3.2 It is common (but often incorrect) to fit an analysis of variance model for the difference post/pre, i.e., posttest2−pretest2 (Model 2) rather than fitting the linear regression as in Model 3. This is equivalent to using an offset, i.e., a covariate with a known coefficient (or $\beta = 1$). Does the data support this hypothesis? [2]

Solution: We can test using a Wald test whether $\mathscr{H}_0 : \beta_{\text{pretest2}} = 1$ in Model 2 versus $\mathscr{H}_0 : \beta_{\text{pretest2}} \neq 1$. The Wald statistic is $W = (0.276 - 1)/0.13 = -5.57$, to be compared to a Student-t distribution with 62 degrees of freedom. We clearly reject the null hypothesis.

3.3 The authors compared different models. Which of the above four models from Listing 3 are nested? [2]

Solution: Models 1–3–4 and 2–3–4 are nested. Models 1 and 2 impose different values of β_{pretest2}, respectively 0 and 1, so cannot be nested.

3.4 Write down the mean for each group of Model 4 and show that Model 3 is a simplification of the latter. Write the null and alternative hypotheses in terms of model parameter and conclude based on the output of Listing 4. [2]

Solution: With parameters in the order they appear in the table, the means are

$$\mathsf{E}(\texttt{posttest2} \mid \cdot) = \begin{cases} \beta_0 + \beta_1 + (\beta_3 + \beta_4)\texttt{pretest2}, & \texttt{group} = \text{DR} \\ \beta_0 + \beta_2 + (\beta_3 + \beta_5)\texttt{pretest2}, & \texttt{group} = \text{DRTA} \\ \beta_0 - \beta_1 - \beta_2 + (\beta_3 - \beta_4 - \beta_5)\texttt{pretest2}, & \texttt{group} = \text{TA}. \end{cases}$$

and we retrieve Model 3 upon setting $\mathscr{H}_0 : \beta_4 = \beta_5 = 0$, meaning the slope is the same for `pretest2` for all three experimental groups. Listing 4 shows the F-test, with a statistic of $F(2,60) = 0.10$ and a p-value of 0.9. There is no evidence of non-parallel slopes.

3.5 The authors computed the marginal mean from Model 3 (Table 7), and the pairwise contrasts, reported in Table 8. Based on these, can you conclude about a ranking as to which treatment is the most effective (if higher scores are better)? [2]

Solution: TA is significantly different from both DR and DRTA (Table 8), and Table 7 indicates that the mean of TA is the highest so this treatment should be recommended.

3.6 If the slopes for `pretest2` for each `group` are not parallel, explain why the comparison using marginal means per experimental groups are hazardous. [2]

> **Solution:** If the slopes are not parallel, then the comparison depends on the value of `pretest2`, whereas otherwise we get the same mean difference regardless of the value of `pretest2` we condition on.

3.7 The authors report Levene's test of homogeneity of variance, [2]

```
> car::leveneTest(rstudent(model3) ~ group,
+                 data = BSJ92,
+                 center = "mean")
```

Listing 5: R call for Levene's test of homogeneity of variance

returns a table with the statistic $F(2,63) = 1.51$, and a p-value of $p = 0.23$. What is the purpose of this test, conclude and explain how this impacts your conclusion, if at all.

> **Solution:** This is an F test for the $K_{ij} = |r_{ij} - \overline{r}_j|$, where r_{ij} is the it externally studentized residual ($i = 1, \ldots, 22$) from group $j = 1, 2, 3$. The purpose of the test is to check that the variance of subgroup is equal or not (homoscedasticity).
>
> Technical aside: under the null hypothesis, if all variates are Student-t with $\nu = n - p - 2$ degrees of freedom, mean zero and standard deviation σ, then K_{ij} is a folded Student-t and it's mean is proportional to $\sigma f(\nu)$. The alternative is that the model was incorrect, and that these variance parameters per group are different.
>
> We get $\mathscr{H}_0 : \sigma_1 = \sigma_2 = \sigma_3$ against the alternative that at least one of DR, DRTA and TA have a different variance. Since the p-value is larger than $\alpha = 0.05$, we fail to reject the null and conclude that there is no evidence of differences in variance per group.

3.8 Given that the `pretest2` and `posttest2` are correlated, does this indicate a violation of the independence assumption? Discuss. [2]

> **Solution:** The analysis is done conditional on `pretest2`, so no problem since the observations for the response are from independent pupils.