22.7 Refer to productivity improvement problem 16.7. The economist also has information on annual productivity improvement in the prior year and wished to use this information as a concomitant variable. The data on the prior year's productivity improvement (X_{ij}) follow.

(a) Obtain the residuals for the covariance model (22.3).

Answer: The covariance model is

$$Y_{ij} = \mu. + \tau_i + \gamma(X_{ij} - \overline{X}..) + \varepsilon_{ij}$$

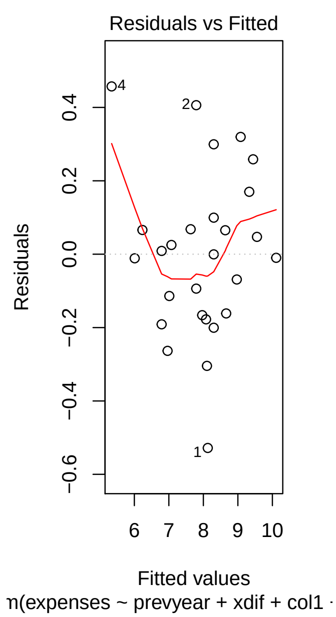
where $\mu.$ is an overall mean, τ_i are the fixed treatment effects subject to the restriction $\sum \tau_i = 0$, γ is a regression coefficient for the relation between Y and X, X_{ij} are constants, ε_{ij} are independent $N(0, \sigma^2)$, and $i = 1, \ldots, r; j = 1, \ldots, n_i$. They sum to zero within the computers accuracy as expected.

Table 1: Residuals

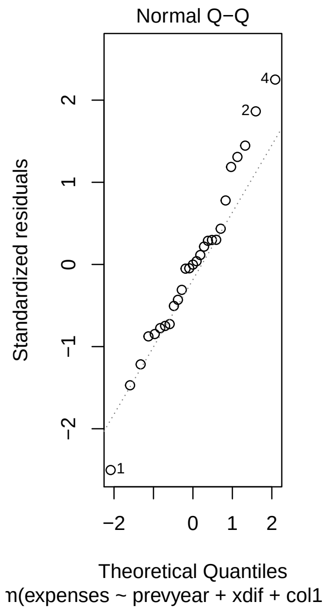
Index $i \downarrow j \rightarrow$	1	2	3	4	5	6	7	8	9	10	11	12
Low	-0.528	0.406	0.0089	0.457	-0.114	-0.191	0.0668	-0.094	-0.011			
Moderate	-0.263	-0.200	0.319	0.299	-0.167	0.068	-0.068	-0.178	0	0.065	0.0251	0.099
high	-0.161	0.259	-0.009	-0.304	0.047	0.17						

(b) For each treatment, plot the residuals against the fitted values. Also prepare a normal probability plot of the residuals and calculate the coefficient of correlation between the ordered residuals and their expected values under normality. What do you conclude from your analysis?

Answer: The correlation is 0.986.



(c) State the generalized regression model to be employed for testing whether or not the treatment regression lines have the same slope. Conduct this test using $\alpha = 0.01$. State the alternatives, decision rule, and conclusion. What is the P-value of the test?



Answer: The reduced model is

$$Y_{ij} = \mu. + \tau_1 I_{ij1} + \tau_2 I_{ij2} + \gamma x_{ij} + \varepsilon_{ij}$$

where

$$I_1 = \begin{cases} 1 & \text{if received treatment 1} \\ -1 & \text{if received treatment 3} \\ 0 & \text{otherwise} \end{cases}$$

and

$$I_2 = \begin{cases} 1 & \text{if received treatment 2} \\ -1 & \text{if received treatment 3} \\ 0 & \text{otherwise} \end{cases}$$

The full model is

$$Y_{ij} = \mu. + \tau_1 I_{ij1} + \tau_2 I_{ij2} + \gamma x_{ij} + \varepsilon_{ij} + \beta_1 I_{ij1} x_{ij} + \beta_2 I_{ij2} x_{ij} + \varepsilon_{ij}$$

The test for parallel slopes is equivalent to testing for no interactions in generalized model

$$H_0 : \beta_1 = \beta_2 = 0$$
$$H_a : \text{not both } \beta_1 \text{ and } \beta_2 \text{ equal zero}$$

Our test statistic is

$$F^* = \frac{\text{SSE}(R) - \text{SSE}(F)}{(n_T - 2) - [n_T - (r+1)]} \nabla \cdot \frac{\text{SSE}(F)}{n_T - (r+1)}$$

In our case, from the ANCOVA tables (which also give us the appropriate degrees of freedom) of the reduced and full model is:

$$\frac{1.3175 - 0.9572}{23 - 21} \nabla \cdot \frac{0.9572}{21} = 3.95$$

We compare this to the F-value with degrees of freedom of first term and second term in denominators of expression calculated above for the F-stat. That is, we compare to $F(1 - .01; 2, 21) = 5.96$

Table 2: Reduced model 22.7

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
prevyear	1	29.97	29.97	523.25	0.0000
col1	1	3.77	3.77	65.87	0.0000
col2	1	0.42	0.42	7.38	0.0123
Residuals	$\boxed{23}$	$\boxed{1.32}$	0.06		

Table 3: Full model 22.7

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
prevyear	1	29.97	29.97	657.61	0.0000
col1	1	3.77	3.77	82.78	0.0000
col2	1	0.42	0.42	9.27	0.0062
col1x	1	0.11	0.11	2.49	0.1299
col2x	1	0.25	0.25	5.42	0.0300
Residuals	21	$\boxed{0.96}$	0.05		

Therefore, we conclude the null. The p-value is 0.033. Therefore, more proof we do not reject at $\alpha = 0.01$.

(d) Could you conduct a formal test here as to whether the regression functions are linear? If so, how many degrees of freedom are there for the denominator mean square in the test statistic?

Answer:

The relevant R-code for this question:

```
library(tidyverse)
library(dplyr)
library(xtable)
data227 = read.table("http://users.stat.ufl.edu/~rrandles/sta4210/Rclassnotes/data/
    textdatasets/KutnerData/Chapter%2022%20Data%20Sets/CH22PR07.txt")
data227$i=data227$V2
data227$j=data227$V3
data227$expenses=data227$V1
data227$prevyear=data227$V4
data227=data227%>%
select(-c(V1,V2,V3,V4))




data227full=data227%>%
mutate(xdif=prevyear-mean(prevyear),
col1=ifelse(i==1,1,
ifelse(i==3,-1,0)),
col2=ifelse(i==2,1,
ifelse(i==3,-1,0)),
col1x=ifelse(i==1,xdif,
ifelse(i==3, xdif, 0)),
col2x=ifelse(i==2,xdif,
ifelse(i==3, xdif, 0)))
data227full
mod=lm(expenses~prevyear+xdif+as.factor(col1)+as.factor(col2),data=data227full)
#part (b)
#mod$residuals
plot(mod,1)
plot(mod,2)

anova(mod)
```

```
mse=anova(mod)[4,3]
N=length(data227$expenses)
ExpVals2 = sapply(1:N, function(k) sqrt(mse) *qnorm((k-.375)/(N+.25)))
cor(ExpVals2, sort(mod$residuals))

#the full model
genmod=lm(expenses~prevyear+xdif+as.factor(col1)+as.factor(col2)+col1x+col2x,data=data227full)
xtable(anova(mod))
anov
Fstar=((1.3174-.9575)/2)/(.9572/21)
fval=qf(.99,2,23)

pval=1-pf(Fstar, 2,23)
```

22.17 Refer to eye contact problem 19.12. Age of personnel officers is to be used as a concomitant variable. The ages (X_{ijk}) of the personnel officers follow.

(a) Obtain the residuals for covariance model (22.26).

Answer: Covariance model (22.26) for a two-factor study with a single concomitant variable, assuming the relation between Y and the concomitant variable X is linear, is:

$$Y_{ijk} = \mu_{..} + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \gamma(X_{ijk} - \overline{X}_{...}) + \varepsilon_{ijk}$$

where $i = 1,\dots,a; j = 1,\dots,b; k = 1,\dots,n$. From a regression perspective, this is equivalent to:

$$Y_{ijk} = \mu_{..} + \alpha_1 I_{ijk1} + \beta_1 I_{ijk2} + (\alpha\beta)_{11} I_{ijk1} I_{ijk2} + \gamma x_{ijk} + \varepsilon_{ijk}$$

where

$$I_1 = \begin{cases} 1 & \text{if case from level for 1 factor A} \\ -1 & \text{if case from level 2 for factor A} \end{cases}$$

and

$$I_2 = \begin{cases} 1 & \text{if case from level 1 for factor B} \\ -1 & \text{if case from level 2 for factor B} \end{cases}$$

and $x_{ijk} = X_{ijk} - \overline{X}_{...}$, where again X_{ijk} are the entries of the concomitant variable (i.e. the covariates, or "exogenous" variables).

Table 4: Residuals for multi-factor study in 19.12

Male	Contact	No contact
1	0.171	-0.816
2	0.081	1.40
3	-0.458	-0.838
4	-1.245	-0.579
5	1.45	0.836

Female	Contact	No contact
1	0.303	0.207
2	1.045	1.578
3	-0.719	-2.097
4	-0.978	0.667
5	0.348	-0.355

They essentially sum to 0.

(b) For each treatment, plot the residuals against the fitted values. Also, prepare a normal probability plot of the residuals and calculate the coefficient of correlation between the ordered residuals and their expected values under normality. What do you conclude from your analysis?

Answer:

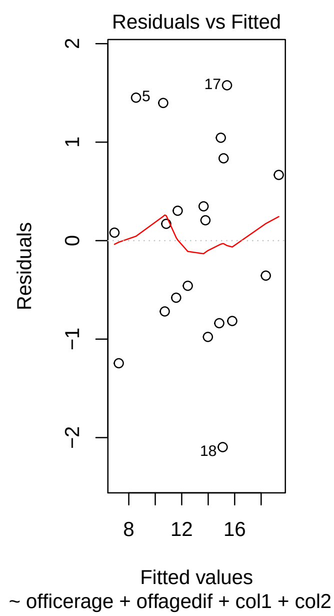


Figure 1: Residuals vs fitted values.

The coefficient of correlation is 0.982.

(c) State the generalized regression model to be employed for testing whether or not the treatment regression lines have the same slope. Conduct this test using $\alpha = 0.005$. State the alternatives, decision rule, and conclusion. What is the P-value of the test?

Answer: The generalized regression model is

$$Y_{ijk} = \mu_{..} + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \gamma(X_{ijk} - \overline{X}_{...}) + \alpha_1' I_{ij1} x_{ij} + \beta_1' I_{ij2} x_{ij} + (\alpha'\beta')_1 \varepsilon_{ijk}$$

The hypothesis is that all $(\alpha\beta)_{ij} = 0$, with the alternative being that not all are zero. Our F-statistic comes from our respective ANCOVA tables again:

Table 5: ANCOVA table Reduced model 22.17

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
offagedif	1	188.39	188.39	152.45	0.0000
gender	1	0.21	0.21	0.17	0.6863
eye contact	1	21.65	21.65	17.52	0.0008
gender:eye contact	1	0.16	0.16	0.13	0.7200
Residuals	15	18.54	1.24		

$$\frac{18.54 - 16.88}{15 - 12} / \frac{16.88}{12} = 0.393$$

We compare this to the F-value with degrees of freedom of first term and second term in denominators of expression calculated above for the F-stat. That is, we compare to $F(1 - .001; 3, 12) = 10.82$

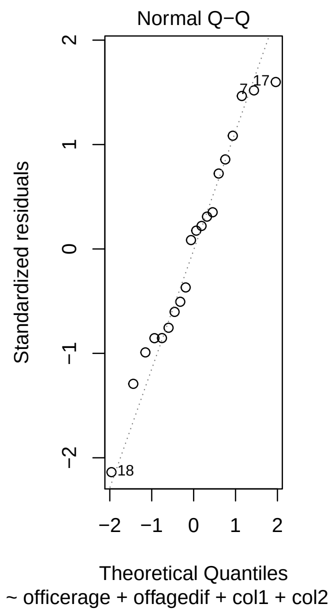


Figure 2: Normal probability plot.

Table 6: ANCOVA table for full model 22.17

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
officeagedif	1	188.39	188.39	133.92	0.0000
gender	1	0.21	0.21	0.15	0.7062
eye contact	1	21.65	21.65	15.39	0.0020
offagedif gender	1	1.20	1.20	0.85	0.3737
offage dif eye econtact	1	0.44	0.44	0.31	0.5873
gender:eye contact	1	0.15	0.15	0.10	0.7529
officerage:gender:eye contact	1	0.04	0.04	0.02	0.8770
Residuals	12	16.88	1.41		

Therefore, we conclude the null. The p-value is 0.760. Therefore, more proof we do not reject at $\alpha = 0.001$.

The R-code for this question:

```
data2217 = read.table("http://users.stat.ufl.edu/~rrandles/sta4210/Rclassnotes/data/
textdatasets/KutnerData/Chapter%2022%20Data%20Sets/CH22PR17.txt")
colnames(data2217)[2]='gender'
colnames(data2217)[3]='eyecontact'
colnames(data2217)[1]='score'
colnames(data2217)[5]='officerage'
data2217full=data2217%>%
mutate(offagedif=officerage-mean(officerage),
#alpha and beta calculated from regression
col1=ifelse(gender==1,1, -1),
col2=ifelse(eyecontact==1,1,-1),
#gamma calculated from regression
col1x=ifelse(gender==1,offagedif,0),
col2x=ifelse(eyecontact==2,offagedif, 0))
data2217
mod2217=lm(score~offagedif+col1+col2+(col1*col2),data=data2217full)
#lm(score~officerage+offagedif+col1+col2+(col1*col2)+col1x+col2x,data2217full)
xtable(anova(mod2217))
```

```
#first 5 are gender==1 and eye contact==1, then gender==1, eye contact==2
#then next factor level is gender==2 and eye contact==1, then gender==2, eye contact==2
mod2217$residuals

mse2=anova(mod2217)[4,3]

N2=length(data2217$score)
ExpVals= sapply(1:N2, function(k) sqrt(mse2) *qnorm((k-.375)/(N+.25)))
cor(ExpVals, sort(mod2217$residuals))

mod2217full=lm(score~offagedif+as.factor(col1)+as.factor(col2)+as.factor(col1*col2)+col1x+
    col2x+offagedif*col1+
offagedif*col2+offagedif*(col1*col2),data2217full)

fstat2=((18.54-16.88)/3)/(16.88/12)
critf=qf(1-.001, 3,12)
pval2=1-pf(fstat2,3,12)
```

22.18 Refer to eye contact problems 19.12 and 22.17. Assume that covariance model (22.26) is applicable.

(a) Test for interaction effects; use $\alpha = 0.01$. State the alternatives, decision rule, and conclusion. What is the P-value of the test?

Answer: Our test is whether or not interaction effect=0. That is

$$H_0 : \text{all } (\alpha\beta)_{ij} = 0$$
$$H_a : \text{not all } (\alpha\beta)_{ij} \text{ equal zero}$$

We can use the ANCOVA table from 22.17 here (the reduced model). The F-stat for the gender-eye contact interaction effect is 0.13 with p-value of 0.72. Clearly, we conclude H_0. The critical value here is $F(1-\alpha; n_{\alpha\beta}, n_{\text{MSE}})$. That is,

$$F(1-0.01;1,15) = 8.68$$

(b) Test for factor A main effects; use $\alpha = 0.01$. State the alternatives, decision rule, and conclusion. What is the P-value of the test?

Answer: Our test is whether or not there are differences in mean across the factor A (gender) . That is

$$H_0 : \mu_{1\cdot} = \mu_{2\cdot}$$
$$H_a : \text{not all } \mu_{i\cdot} \text{ are equal}$$

or equivalently

$$H_0 : \alpha_1 = \alpha_2 = 0$$
$$H_a : \text{not all } \alpha_i \text{ are equal}$$

From the table: the F-value for gender effects is 0.17 and the p-value is 0.683. We conclude H_0.

(c) Test for factor B main effects; use $\alpha = 0.01$. State the alternatives, decision rule, and conclusion. What is the P-value of the test?

Answer: Our test is whether or not there are differences in mean across the factor B (eye contact) . That is

$$H_0 : \mu_{1\cdot} = \mu_{2\cdot}$$
$$H_a : \text{not all } \mu_{i\cdot} \text{ are equal}$$

or equivalently

$$H_0 : \beta_1 = \beta_2 = 0$$
$$H_a : \text{not all } \beta_i. \text{ are equal}$$

We find our $F^* = 17.52$ with p-value of 0.0008. The critical value is $F(0.99; 1, 15) = 8.68$. We conclude H_a.

(d) Compare the gender main effects by means of a 99 percent confidence interval. Interpret your interval estimate.

Answer:

$$12.96 - .29I_1 - 1.114I_2 + 0.104I_1I_2 + 0.402x$$

So

$$\hat{D} = \hat{\beta}_1 - \hat{\beta}_2 = 2\hat{\beta}_1 = 2 \cdot -1.114 = -2.23$$

(e) Estimate the mean success rating by female personnel officers aged 40 when eye contact is present; use a 99 percent confidence interval.

Answer:

23.6 Refer to cash offers problem 19.10. Suppose that observations $Y_{214} = 28$ and $Y_{323} = 20$ are missing because the offer received in each of these cases was a trade-in offer, not a cash offer.

(a) State the ANOVA model for this case. Also state the equivalent regression model; use 1,-1,0 indicator variables.

Answer: Define A is the gender effect and B is the age effect. The model is developed from the factor effects model (19.23)

$$Y_{ijk} = \mu_{..} + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \varepsilon_{ijk}$$

The regression model equivalent is

$$Y_{ijk} = \mu_{..} + \alpha_1 X_{ijk1} + \beta_1 X_{ijk2} + \beta_2 X_{ijk3} + (\alpha\beta)_{11} X_{ijk1} X_{ijk2} + (\alpha\beta)_{12} X_{ijk1} X_{ijk3} + \varepsilon ijk$$

Where

$$X_1 = \begin{cases} 1 & \text{if case from level 1 for factor A} \\ -1 & \text{if case from level 2 for factor A} \end{cases}$$

$$X_2 = \begin{cases} 1 & \text{if case from level 1 for factor B} \\ -1 & \text{if case from level 3 for factor B} \\ 0 & \text{otherwise} \end{cases}$$

$$X_3 = \begin{cases} 1 & \text{if case from level 2 for factor B} \\ -1 & \text{if case from level 3 for factor B} \\ 0 & \text{otherwise} \end{cases}$$

The α_1 represents the A main effect, the β_1 and β_2 terms are for the B main effect, and the next two represent the AB interaction effect.

(b) What is the reduced regression model for testing for interaction effects?

Answer: The reduced regression model is

$$Y_{ijk} = \mu_{..} + \alpha_1 X_{ijk1} + \beta_1 X_{ijk2} + \beta_2 X_{ijk3} + \varepsilon_{ijk}$$

Table 7: ANOVA reduced 23.6

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
gender1	1	1.06	1.06	0.42	0.5216
age2	1	0.24	0.24	0.10	0.7590
age3	1	289.65	289.65	115.06	0.0000
Residuals	30	75.52	2.52		

Table 8: ANOVA full 23.6

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
gender1	1	1.06	1.06	0.42	0.5244
age2	1	0.24	0.24	0.09	0.7605
age3	1	289.65	289.65	113.69	0.0000
gender1:age2	1	2.69	2.69	1.06	0.3127
gender1:age3	1	1.50	1.50	0.59	0.4500
Residuals	28	71.33	2.55		

(c) Test whether or not interaction effects are present by fitting the full and reduced regression models; use $\alpha = 0.05$. State the alternatives, decision rule, and conclusion. What is the p-value of the test?

Answer: The reduced model ANOVA table and the full: Our test is whether or not interaction effect=0. That is

$$H_0 : \text{all } (\alpha\beta)_{ij} = 0$$
$$H_a : \text{not all } (\alpha\beta)_{ij} \text{ equal zero}$$

We find from the table

$$F^* = \frac{75.52 - 71.33}{30 - 28} / \frac{71.33}{28} = 0.822$$

Compare this to

$$F(1-\alpha, n_R - n_F, n_F) = 3.31$$

So we conclude H_0 at the specified confidence coefficient. The p-value is 0.449.

(d) State the reduced regression models for testing for age and gender main effects, respectively, and conduct each of the tests. Use $\alpha = 0.05$ each time and state the alternatives, decision rule, and conclusion. What is the p-value of each test?

Answer: For gender effects we test:

$$H_0 : \alpha_1 = 0$$
$$H_a : \alpha_1 \neq 0$$

From the ANOVA table, the full one in this case, For age effects we test

$$H_0 : \beta_1 = \beta_2 = 0$$
$$H_a : \text{not all } \alpha_i \text{ are equal}$$

The reduced regression models for testing for factor A main effects and factor B main effects are

$$Y_{ijk} = \mu_{..} + \beta_1 X_{ijk2} + \beta_2 X_{ijk3} + (\alpha\beta)_{11} X_{ijk1} X_{ijk2} + (\alpha\beta)_{12} X_{ijk1} X_{ijk3} + \varepsilon_{ijk}$$

This leads to SSE of 75.871 with 29 degrees of freedom. We compare this to our full model from earlier in the same way.

$$F^* = \frac{75.81 - 71.33}{29 - 28} / \frac{71.33}{28} = 1.76$$

Compare to critical value of $F(.95,1,28)=4.18$ as in part (b), we therefore conclude H_0. The p-value is 0.195. and for factor B main effects are:

$$Y_{ijk}=\mu_{..}+\alpha_1 X_{ijk1}+(\alpha\beta)_{11}X_{ijk1}X_{ijk2}+(\alpha\beta)_{12}X_{ijk1}X_{ijk3}+\varepsilon_{ijk}$$

This leads to SSE of 359.94 with 30 dof. We compare this to our full model from earlier in the same way as in part (c).

$$F^*=\frac{359.94-71.33}{30-28}/\frac{71.33}{28}=56.65$$

Compare to 3.31 as in part (b). Conclude H_a that we have age effects. The p-value is 0 essentially.

(e) To study the nature of the age main effects, estimate the following pairwise comparisons:

$$D_1=\mu_{1.}-\mu_{2.} \qquad D_2=\mu_{1.}-\mu_{3.} \qquad D_3=\mu_{2.}-\mu_{3.}$$

Use the most efficient multiple comparison procedure with a 90 percent family confidence coefficient.

Answer:

$$\hat{D}_1=21.5-27.727=-6.22$$
$$\hat{D}_2=21.5-21.545=-0.045$$
$$\hat{D}_3=27.727-21.545=6.18$$

We note $s^2(\hat{D})=\frac{\text{MSE}}{a^2}\sum_j\left(\frac{1}{n_{ij}+n_{i'j}}\right)$ for the age effects.

In our example, that means

$$s^2(\hat{D}_1)=s^2(\hat{D}_2)=0.668 \qquad s^2(\hat{D}_3)=0.684$$

We check the most efficient methods (where a=3):

$$B=t(1-.1/(2\cdot 3),28)=2.24$$
$$S=\sqrt{((b-1)F(1-.1,b-1,28)}=2.24$$
$$T=\frac{1}{\sqrt{2}}q(1-.1,28)=2.14$$

So we use Tukey multiple. Therefore, we have

$$-6.22\pm 2.14\cdot 0.667\longrightarrow -7.66\le D_1\le -4.805$$
$$-0.045\pm 2.14\cdot 0.667\longrightarrow -1.395\le D_2\le 1.462$$
$$6.18\pm 2.14\cdot(0.683)\longrightarrow 4.80\le D_3\le 7.73$$

(f) In the population of female owners, 30 percent are young, 60 percent are middle aged, and 10 percent are elderly. Estimate the mean cash offer for this population with a 95 percent confidence interval.

Answer: We note

$$\hat{L}=0.3\cdot\overline{Y}_{12}+0.6\cdot\overline{Y}_{22}+0.1\cdot\overline{Y}_{32}$$

and plugging in numbers yield:

$$\hat{L}=0.3*21.33+0.6*27.66+0.1*20.5=25.05$$

We note:

$$s^2(\hat{L})=\frac{\text{MSE}}{b^2}\sum_i c_i^2\sum_j\frac{1}{n_{ij}}$$

where we calculated MSE to be 2.548 from the full model. We find $s(\hat{L}) = 0.443$. For a 95 percent confidence interval, as usual $t(1-\alpha/2, 28) = 2.048$. Therefore, our interval is

$$25.05 \pm 0.443 \cdot 2.048 \longrightarrow 24.14 \leq L \leq 25.96$$

Here is some "code".

```
data236 = read.table("http://users.stat.ufl.edu/~rrandles/sta4210/Rclassnotes/data/
     textdatasets/KutnerData/Chapter%2019%20Data%20Sets/CH19PR10.txt")
colnames(data236)=c('cash','age', 'gender', 'entry')
#remove columns (manually)
data236=data236[c(-16,-33),]

data236full=data236%>%
mutate(gender1=ifelse(gender==1, 1,-1),
age2=ifelse(age==1, 1,
ifelse(age==3, -1,0)),
age3=ifelse(age==2, 1,
ifelse(age==3, -1, 0)),
gender1age2=gender1*age2,
gender1age3=gender1*age3)
data236full%>%
filter(c(age==2&gender==2))
mod236=lm(cash~gender1+age2+age3,data=data236full)
xtable(anova(mod236full))

mod236full=lm(cash~gender1+age2+age3+gender1*age2+gender1*age3, data=data236full)
anova(mod236full)*1
intefftestval=((75.52-71.33)/2)/(71.33/28)
fcrit=qf(0.95, 2,28)
1-pf(intefftestval, 2,28)
#in R have to separatetely do calculation for
#genderage interaction * tells R to do full model!
#doesnt matter above because that is what we want
effmodage=lm(cash~gender1+gender1age2+gender1age3, data=data236full)
anova(effmodage)
#check the fstat reduced vs full((359.94-71.33)/2)/(71.33/28)

effmodgender=lm(cash~age2+age3+gender1age2+gender1age3, data=data236full)
anova(effmodgender)

#linear test is reduced vs full((75.81-71.33)/1)/(71.33/29)
medif1=mean(data236$cash[data236$age==1])-mean(data236$cash[data236$age==2])
medif2=mean(data236$cash[data236$age==1])-mean(data236$cash[data236$age==3])
medif3=mean(data236$cash[data236$age==2])-mean(data236$cash[data236$age==3])

a=length(unique(data236$gender))
b=length(unique(data236$age))
n1j=length(data236$gender[data236$gender==1])

n11=1/length(data236$cash[data236$gender==1&data236$age==1])
n12=1/length(data236$cash[data236$gender==1&data236$age==2])
n13=1/length(data236$cash[data236$gender==1&data236$age==3])

n21=1/length(data236$cash[data236$gender==2&data236$age==1])
n22=1/length(data236$cash[data236$gender==2&data236$age==2])
n23=1/length(data236$cash[data236$gender==2&data236$age==3])

sD1=sqrt(2.55/(a^2)*(n11+n12+n21+n22))
sD3=sqrt(2.55*(1/(a^2))*(1/5+1/6+1/6+1/5))#(n12+n13+n22+n23))

inner=(.3^2+.6^2+.1^2)
sdnew=sqrt(2.55*(.3^2*n11+.6^2*n13+.1^2*n23))
25.05+sdnew*qt(0.975, 28)
```

23.20 Refer to problem 19.10. It is known that in both populations of male and female owners, 30 percent are young, 60 percent are middle aged, and 10 percent are elderly. Test by means of the single degree of freedom t^* test statistic whether or not the mean cash offers for male and female owners are equal; use $\alpha = 0.05$. State the alternatives, decision rule and conclusion. What is the p-value of the test?

Answer: We must take the weighted mean for each gender. i.e. for male (gender==1) (recalling we made factor A gender and factor B age)

$$L_1 = 0.3 \cdot \mu_{11} + 0.6 \cdot \mu_{12} + 0.1 \cdot \mu_{13}$$

and for female:

$$L_2 = 0.3 \cdot \mu_{21} + 0.6 \cdot \mu_{22} + 0.1 \cdot \mu_{23}$$

We find:

$$\hat{L}_1 = 0.3 * 21.66 + 0.6 * 27.83 + 0.1 * 22.33 = 25.433$$
$$\hat{L}_2 = 0.3 * 21.33 + 0.6 * 27.66 + 0.1 * 20.5 = 25.05$$

Our hypothesis test is that $H_0 : L1 - L2 = 0$ with alternative $L_1 \neq L_2$. The difference is

$$\hat{L}_1 - \hat{L}_2 = 0.383$$

$$s^2(\hat{L}) = \frac{\text{MSE}}{a \cdot b} \sum_i c_i^2 \sum_j \frac{1}{n_{ij}} \to s(\hat{L}) = 0.602$$

We find the MSE from fitting the full model same as in 23.6 but this time we do not remove the data points. This leads to an MSE of 2.39, as seen in the ANOVA table.

Table 9: ANOVA table 23.20, full model no removed data points

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
gender1	1	5.44	5.44	2.28	0.1416
age2	1	0.04	0.04	0.02	0.8958
age3	1	316.68	316.68	132.56	0.0000
gender1:age2	1	3.37	3.37	1.41	0.2439
gender1:age3	1	1.68	1.68	0.70	0.4082
Residuals	30	71.67	2.39		

Our $t^* = \frac{\hat{L}_1 - \hat{L}_2}{s(\hat{L})} = \frac{0.383}{0.602} = 0.63$. We can compare this to $t(1 - \alpha/2; n_T - 2) = 2.04$. Therefore, if $|t^*| \leq 2.04$, we conclude H_0. Therefore, we conclude H_0.