第一大题：协方差分析：

***Methods and assumption checks:***

数据特征 —— 为什么用此类型分析：have two explanatory variables, a grouping explanatory variable with two levels and a numeric explanatory variable, so have fitted a linear model with both variables and included an interaction term.

为什么对因变量取log：As the (response variable) increased the variabilty also increased so we logged the (response variable) data, this evened out the scatter, makes our model fit better.

套话：The test for the interaction term proved to be (significant / not significant), so the interaction term was (kept / not kept) and the model could not be simplified further.

The residuals were fine, there were no problems with normality and no unduly influential points. We have independence from taking a random sample.

Our model is:

`5[]1F@2NQ5$}VEC6PORYS2

8X)FUAIL~94RZ4J)QS}JV[G

如何写某一个level的回归模型：

根据上述模型，

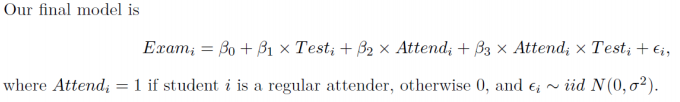
病人：

log(Ventricles\_i) = beta0 + beta1 x Age\_i

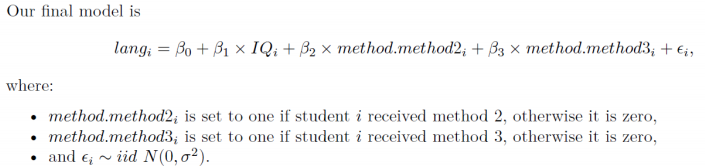
健康人：

log(Ventricles\_i) = (beta0 + beta2) + (beta1 + beta3) x Age\_i

或didn’t log the response variable的形式：



或a grouping explanatory variable with more than two levels：



Our model explained xx% of the variability in the data.

***Executive Summary:***

Some questions:

1.In terms of slopes and/or intercepts, explain what the coefficient of Age:ADH is estimating. （解释interaction item的含义）：

The coefficient is estimating the difference in slope between the line for Age versus log(Ventricles) when people have dementia symptoms (the base line) and the slope of the line for Age versus log(Ventricles) for healthy people.

ADH对Age对log(Ventricles)的影响的影响。

2.问什么答什么，注意是点估计还是区间估计

第二大题：双因素方差分析：

***Methods and assumption checks:***

数据特征 —— 为什么用此类型分析：We have two explanatory factors and a continuous response variable, so have fitted a Twoway ANOVA model to the data.

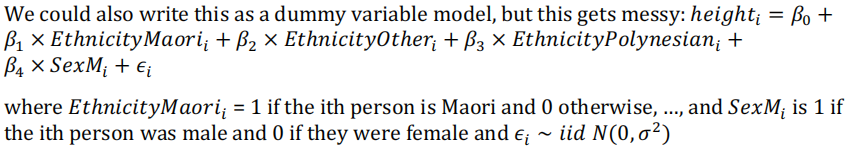
套话：The test for the interaction term proved to be (significant / not significant), so the interaction term was (kept / not kept) and the model could not be simplified further.

The residuals were fine, there were no problems with normality and no unduly influential points. We have independence from taking a random sample.

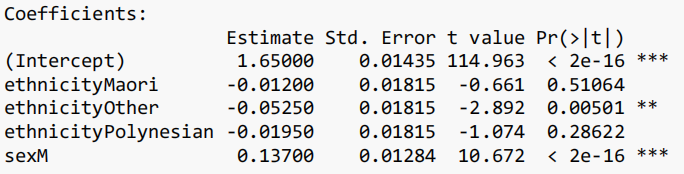
Our model is:

IMG_256

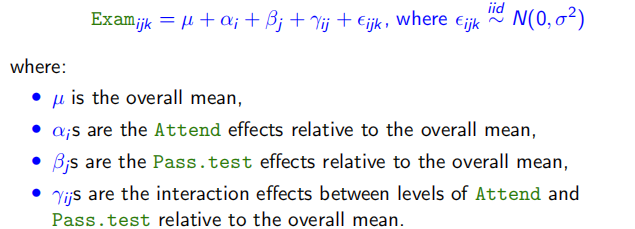
IMG_256

或

（由于类别中有多个组别，不能再单纯用Ethnicity和Sex作为因变量，可从summary函数的结果中读出有什么变量）

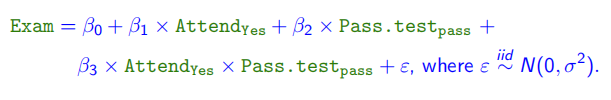


we have a statistically significant interaction：



（r作为交互项添加到模型中）

或：

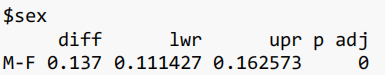


Our model explained xx% of the variability in the data.

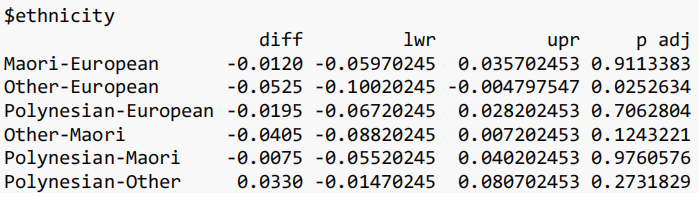
***Executive Summary:***

We estimate that, on average, males are between 11cm and 16cm taller than females.

summary2way(height.fit2,page="nointeraction")：



There was little evidence of differences in height between the four ethnic groups. The only significant difference found was that Europeans were taller than the “Other” ethnicities group. We estimate that, on average, Europeans are between 0.5cm and 10cm taller than “Other” ethnicities.



（只找p<0.05的项，并解释其含义）

为什么summary2way()计算得到的数据项p值比summary()大？

summary2way is more rough, requiring all data items to fall within the 95% confidence interval.

第三大题：广义线性模型（泊松回归）

***Methods and assumption checks:***

数据特征 —— 为什么用此类型分析：The xxxx as the response variable is a count variable, so we choose the Poisson GLM. The explanatory variables were xxxx and xxxx and......

套话：

Residuals looked (fine / not good), there was evidence that over–dispersion（overdispersed） was present (P-value = 0.0457) so the model was refitted as quasipoisson.

The test for the interaction term proved to be (significant / not significant), so the interaction term was (kept / not kept) and the model could not be simplified further.

（一定要先确实模型，再讨论交互项的显著性）

The final model is：

IMG_256

where μi is the expected number of xxxx. Yi（响应变量） is the number of xxxx and has an overdispersed distribution with mean μi.

或：log(μi) = beta0 + beta1 × yeari

***Executive Summary:***

1.We are interested in.......

2.We found strong evidence suggesting there is a (positive/negative) relationship between xx and xx (P-value < 0.05).

3(定量描述均值、C.I.). 与普通线性模型不同的是，描述时要使用log取exp变换之后的值，不用直接使用log变换后的值。

For each additional 1 dollar/1 age/.... increase in xx, the average xx increased by somewhere between xx% and xx%.

β1 > 0, median increase by（增加了多少，(exp(β1)-1) x 100%）

β1 < 0, median decrease by（降低了多少, (1-exp(β1)) x 100%）

第四大题：广义线性模型（逻辑回归）

***Methods and assumption checks:***

数据特征 —— 为什么用此类型分析：

We have a response variable with two outcomes, survived or died（Yes or No）, so have fitted a logistic regression model. We have (categorical (grouping) / numerical) explanatory variables.

套话：

The test for the interaction term proved to be (significant / not significant), so the interaction term was (kept / not kept) and the model could not be simplified further.

The check for residual deviance has a P-value of 0.168 and hence the Binomial assumption is adequate.

The residual plot shows no problem.

Our model is:

IMG_256

where the odds are the odds of a mouse with given value of explanatory variables surviving. Doselow\_i is 1 if a mouse in the ith group got a low dose of radiation and 0 otherwise, and TreatmentStrept\_i is 0 if a mouse was in the control group and 1 if the mouse was in the group treated with Streptomycin.

（一定要写成Odds，要会解释各变量含义）

***Executive Summary:***

1.We are interested in.......

2.We found strong evidence suggesting there is a (positive/negative) relationship between xx and xx (P-value < 0.05).

3.注意逻辑回归研究的是满足二项分布的自变量的odds，即几率。

Mice treated with streptomycin are more likely to survive. For mice with a given dose of radiation, we estimate the odds that a mouse that was treated with streptomycin survived the first 10 days after exposure are between 3.3 and 7.3 times the odds that a mouse that wasn’t treated survived.