Inferential Statistics with R

Example Data Set

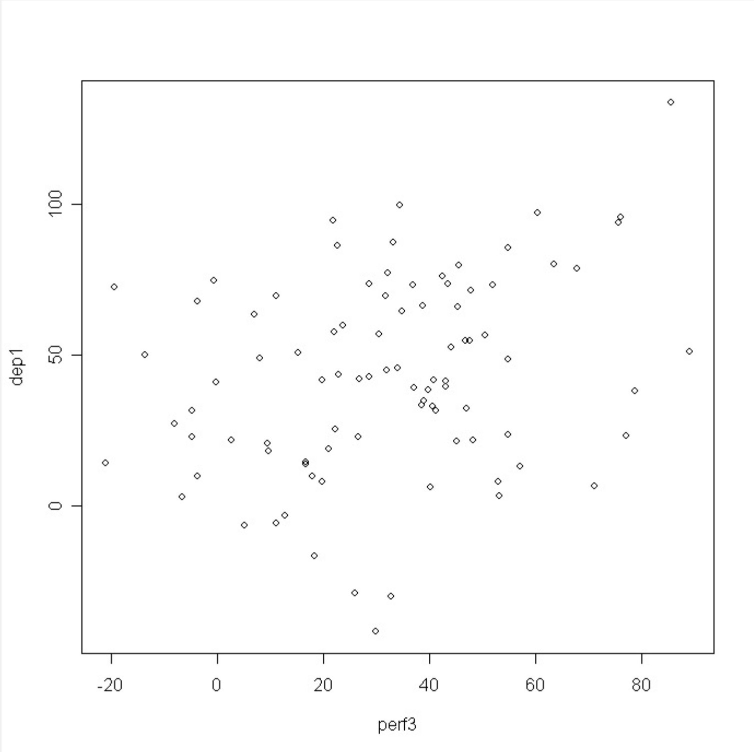
- A researcher is interested in evaluating two therapies for perfectionism; specifically investigating whether they will be effective in reducing levels of perfectionism
  - Levels of perfectionism are recorded at baseline, 1 month (mid intervention) and 2 months (post intervention) for each experimental group (CBT, General Stress) and a control group
- The researcher also records levels of anxiety and depression at each time point, as well as the sex of the subject

Correlation

- Hypothesis #1: Are baseline depression and perfecionism scores correlated?
  - >cor.test(dep1,perf1)
    - Pearson's product-moment correlation
    - data: perf1 and dep1
    - t = 5.2198, df = 88, p-value = 1.183e-06
    - alternative hypothesis: true correlation is not equal to 0
    - 95 percent confidence interval: 0.3103911 0.6298929
    - sample estimates:
    - cor
    - 0.4862279
- Spearman’s rank based correlation coefficient can be computed with:
    - >cor.test(dep1,perf1method=”spearman”)

Simple Regression

- Hypothesis #2: Can we predict posttest perfectionism scores from pretest depression scores?
- Scatterplot
  - >plot(perf3,dep1)



Simple Regression, cont'd

- Create a linear model object and print a summary of the results
  - >mod1<- lm(depres3~anx1)
  - >summary(mod1)
    - Call: lm(formula = perf3 ~ dep1)
    - Coefficients:

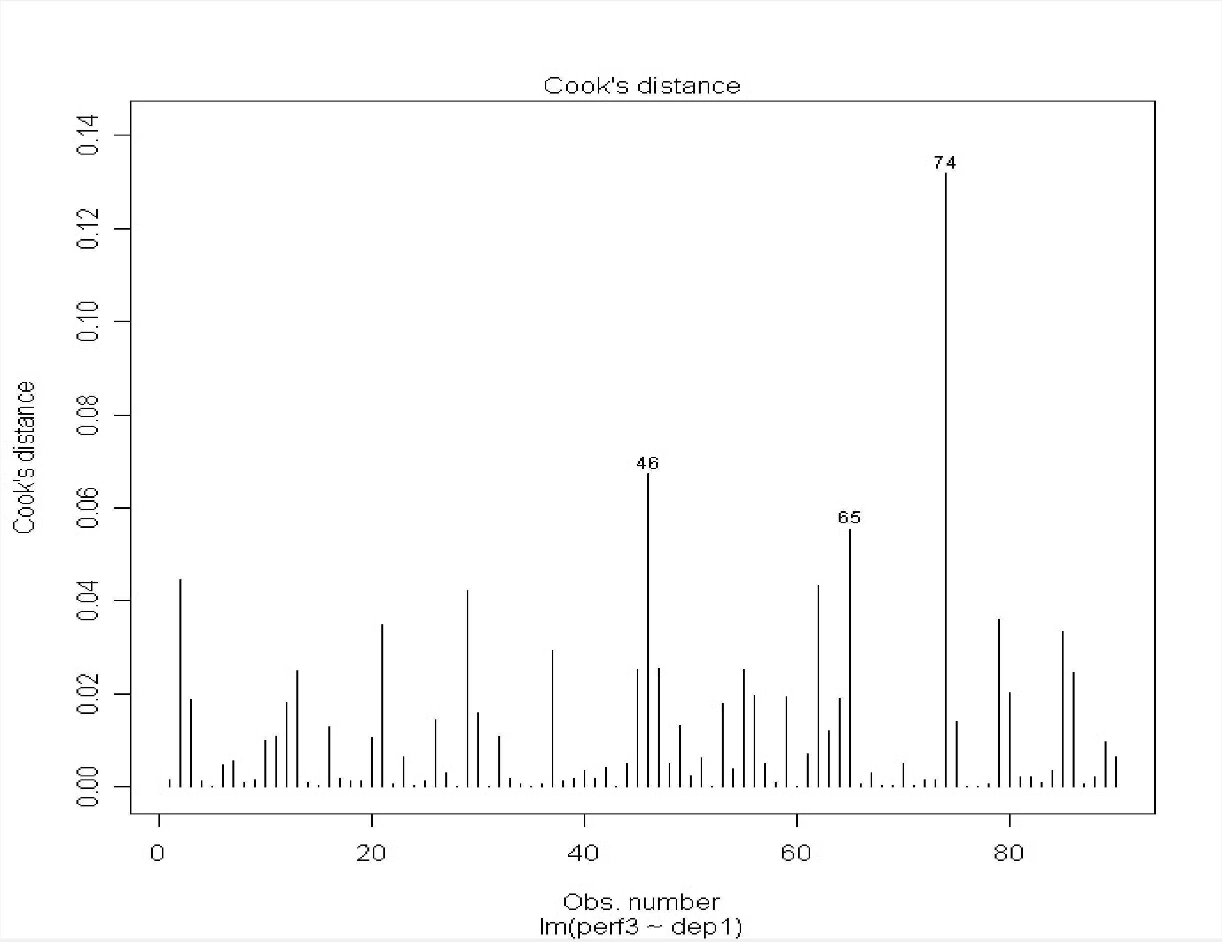
	Estimate	Std. Error	t value	Pr(>|t|)	
(Intercept)	22.8801	4.0309	5.676	1.74e-07	***
dep1	0.2146	0.0748	2.869	0.00515	**

    - Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
    - Residual standard error: 22.98 on 88 df
    - Multiple R-squared: 0.085, Adjusted R-squared: 0.075
    - F-statistic: 8.232 on 1 and 88 DF, p-value: 0.005153

Simple Regression, cont’d

- Regression Diagnostics
  - > install.packages(”car”)
  - > library(car)
  - > outlier.test(mod1)
    - max|rstudent| = 2.626686, degrees of freedom = 87,
    - unadjusted p = 0.01018786, Bonferroni p = 0.9169073
    - Observation: 46
- Other diagnostics are available for identifying normality issues, linearity issues, etc.
  - plot(mod1) provides 4 diagnostic plots
  - Identify influential observations with Cook’s D
    - >cutoff <- 4/((length(perf1)-length(mod1$coefficients)-1))
    - >plot(mod1, which=4, cook.levels=cutoff)

Regression Diagnostics: Identifying Influential Observations

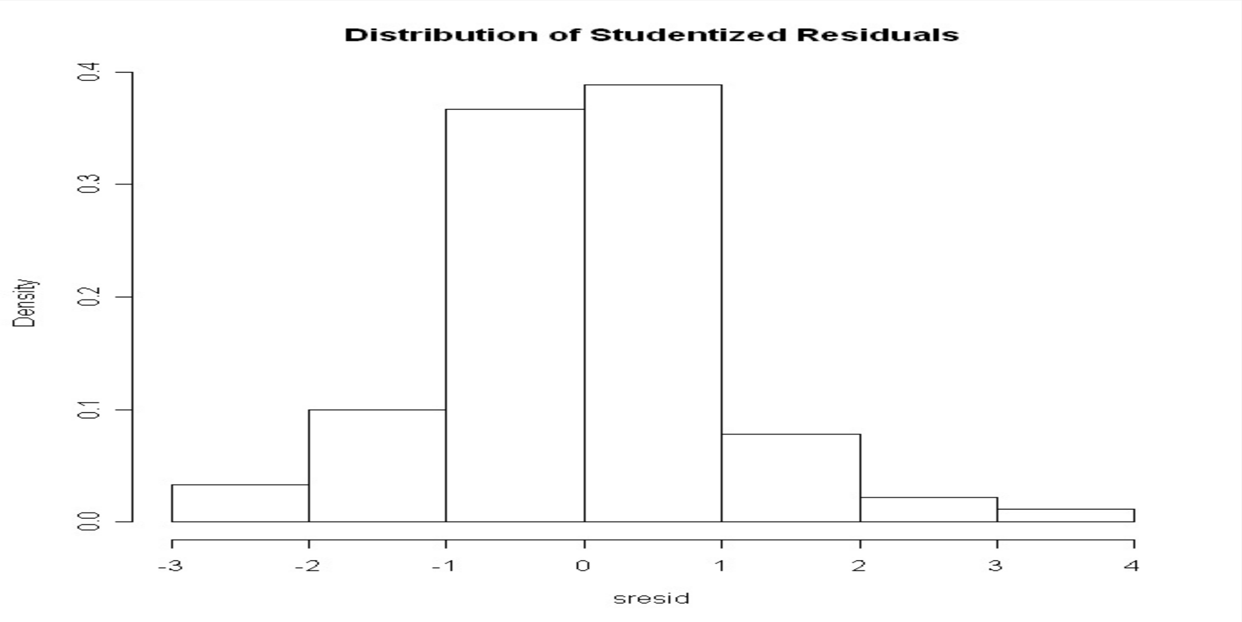


Multiple Regression

- Hypothesis #3: Can posttest perfectionism scores be predicted from depression scores, controlling for pretest perfectionism?
  - >mod2<-lm(perf3 ~ dep1 + perf1)
  - >summary(mod2)
    - Call: lm(formula = perf3 ~ dep1 + perf1)
    - Estimate Std. Error t value Pr(>|t|)
    - (Intercept) -11.84151 4.86248 -2.435 0.0169 *
    - dep1 -0.05558 0.06215 -0.894 0.3736
    - perf1 1.12513 0.12583 8.942 5.92e-14 ***
    - Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
    - Residual standard error: 16.68 on 87 df
    - Multiple R-squared: 0.5235, Adjusted R-squared: 0.5125
    - F-stat: 47.78 on 2 & 87 DF, p-val: 9.933e-15

Multiple Regression, cont'd

- Model Diagnostics
  - Residual Normality
    - >library(MASS)
    - >sresid <- studres(mod2)
    - >hist(sresid, freq=FALSE, main="Distribution of Studentized Residuals")



Multiple Regression, cont'd

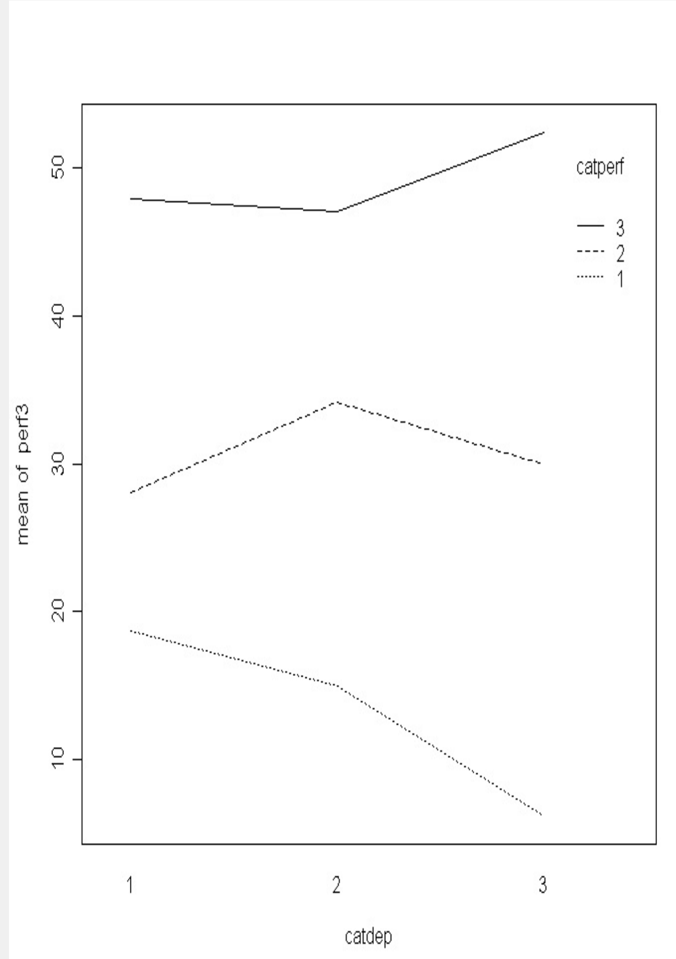
- Adding an interaction term to the model
  - >dep1c<-dep1-mean(dep1)
  - >perf1c<-perf1-mean(perf1)
  - >Mod4<-lm(perf3~dep1c + perf1c + dep1c|perf1c)
  - Or equivalently: >Mod4<-lm(perf3 ~ dep1c*perf1c)
  - >Summary(Mod4)
    - Coefficients:

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	31.143941	1.925312	16.176	< 2e-16 ***
dep1c	-0.059519	0.062046	-0.959	0.340
perf1c	1.101094	0.126964	8.673	2.28e-13 ***
dep1c:perf1c	0.003895	0.003159	1.233	0.221

    - Multiple R-squared: 0.5317, Adjusted R-squared: 0.5154
    - F-stat: 32.55 on 3 & 86 DF, p-value: 3.724e-14

Multiple Regression, cont'd

- Plotting the interaction
  - quantile(dep1c, probs=c(.33, .66))
  - catdep<-1:90
  - catdep[dep1c<=-17]<-1
  - catdep[dep1c>=-17 & dep1c<=14]<-2
  - catdep[dep1c>=14]<-3
  - quantile(perf1c, probs=c(.33, .66))
  - catperf<-1:90
  - catperf[perf1c<=-6]<-1
  - catperf[perf1c>=-6 & perf1c<=5]<-2
  - catperf[perf1c>=5]<-3
  - catdep<-as.factor(catdep)
  - catperf<-as.factor(catperf)
  - interaction.plot(catdep,catperf,perf3)



Independent Samples t-tests

- Hypothesis #4: Is there a difference between males and females on pretest perfectionism?
  - Independent Samples t-test (assuming equal variances)
  - >t.test(perf1[sex == 1],perf1[sex == 2], var.equal=T)
    - Two Sample t-test
    - data: perf1[sex == 1] and perf1[sex == 2]
    - t = 0.81, df = 88, p-value = 0.4201
    - alternative hypothesis: true difference in means is not equal to 0
    - 95 percent confidence interval: -4.101780 9.746436
    - sample estimates:
    - mean of x mean of y
    - 42.92910 40.10677

Independent Samples t-tests Under Variance Heterogeneity

- Welch's two independent samples t-test (not assuming equal variances)
  - >t.test(perf1[sex == 1],perf1[sex == 2])
    - Welch Two Sample t-test
    - data: perf1[sex == 1] and perf1[sex == 2]
    - t = 0.806, df = 71.335, p-value = 0.4229
    - alternative hypothesis: true difference in means is not equal to 0
    - 95 percent confidence interval: -4.159457 9.804113
    - sample estimates:
    - mean of x mean of y
    - 42.92910 40.10677

Independent Samples t-tests Under Nonnormality

- Wilcoxon-Mann-Whitney nonparametric two independent samples test
  - >wilcox.test(perf1[sex == 1],perf1[sex == 2])
    - Wilcoxon rank sum test with continuity correction
    - data: perf1[sex == 1] and perf1[sex == 2]
    - W = 1090, p-value = 0.2932
    - alternative hypothesis: true location shift is not equal to 0
- Which is equivalent to:
    - > rperf<-rank(perf1)
    - > t.test(rperf[sex == 1],rperf[sex == 2], var.equal=T)
      - Two Sample t-test
      - data: rperf[sex == 1] and rperf[sex == 2]
      - t = 1.056, df = 88, p-value = 0.2939
      - alternative hypothesis: true difference in means is not equal to 0

What about a t-test for Nonnormality and Variance Inequality?

- Several procedures have been proposed, although the Welch t-test on trimmed means has far garnered the most attention
  - Problem: there is no built in function for computing the trimmed Welch t in R
  - Solution: Rand Wilcox has written functions that accompany his texts on robust statistics that includes a function for computing the trimmed Welch t (which was develop by Yuen and often referred to as the Yuen test)

Paired Samples t-tests

- Hypothesis #5: Is there a difference between pre and post perfectionism scores?
  - t.test(perf1, perf3, paired=T)
    - Paired t-test
    - data: perf1 and perf3
    - t = 5.1863, df = 89, p-value = 1.334e-06
    - alternative hypothesis: true difference in means is not equal to 0
    - 95 percent confidence interval:
    - 5.601134 12.558353
    - sample estimates:
    - mean of the differences
      - 9.079743

Paired Samples under Nonnormality

- If the distribution of difference scores is not normally distributed, the Wilcoxon signed ranks test can be much more powerful than the paired samples t-test
  - wilcox.test(perf1, perf3, paired=T)
    - Wilcoxon signed rank test with continuity correction
    - data: perf1 and perf3
    - V = 3326, p-value = 2.714e-07
    - alternative hypothesis: true location shift is not equal to 0

One-way Independent Groups ANOVA

- Hypothesis #6: Is there a difference between the three treatment conditions on posttest-perfectionism?
  - > group<-as.factor(group)
  - Option 1:
    - > mod3<- lm(perf3 ~ group)
    - > anova (mod3)
  - Option 2:
    - > mod3<-aov(perf3 ~ group)
    - > summary(mod3)
  - Option 3:
    - >oneway.test(perf3 ~ group, var.equal=T)

One-way Independent Groups ANOVA

- oneway.test(perf3 ~ group, var.equal=T)
  - One-way analysis of means
  - data: perf3 and group
  - F = 2.7896, num df = 2, denom df = 87, p-value = 0.06695
- mod3<- lm(perf3 ~ group)
- anova (mod3)
  - Analysis of Variance Table
  - Response: perf3
  - Df Sum Sq Mean Sq F value Pr(>F)
  - group 2 3062 1531 2.7896 0.06695 .
  - Residuals 87 47750 549
  - Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Multiple Comparisons for One-way Independent Groups ANOVA

- Tukey's Honestly Significant Difference (HSD) Familywise Error Controlling Procedure for Pairwise Comparisons
  - mod4<-aov(perf3 ~ group)
  - TukeyHSD(mod4)
    - Tukey multiple comparisons of means
    - 95% family-wise confidence level
    - Fit: aov(formula = perf3 ~ group)
    - $group
    - diff lwr upr p adj
    - 2-1 -11.033877 -25.457488 3.389734 0.1677661
    - 3-1 2.344296 -12.079315 16.767906 0.9206245
    - 3-2 13.378173 -1.045438 27.801784 0.0748159

Multiple Comparisons for One-way Independent Groups ANOVA

- Flexible procedure for all pairwise comparisons
  - pairwise.t.test(perf3, group, p.adj=”none”)
    - Pairwise comparisons using t tests with pooled SD
    - data: perf3 and group
    -
    -           1   2
    - 2 0.072   -
    - 3 0.699 0.030
    -
    - P value adjustment method: none

Multiple Comparisons for One-way Independent Groups ANOVA

- Multiplicity control with pairwise.t.test
  - pairwise.t.test(perf3, group, p.adjust.method = “bonferroni”)
  - pairwise.t.test(perf3, group, p.adjust.method = “holm”)
  - pairwise.t.test(perf3, group, p.adjust.method = “fdr”)
    - Pairwise comparisons using t tests with pooled SD
    - data: perf3 and group
    - 1 2
    - 2 0.107 -
    - 3 0.699 0.089
    - P value adjustment method: fdr

Multiple Comparisons for One-way Independent Groups ANOVA

- A not efficient way to conduct pairwise comparisons, but that demonstrates the extraction of objects
  - pvals<-(c(t.test(perf3[group==1],perf3[group==2])$p.value, t.test(perf3[group==1],perf3[group==3])$p.value, t.test(perf3[group==2],perf3[group==3])$p.value)
  - p.adjust(pvals, method=”hommell”)
  - p.adjust(pvals, method="hommel")
    - [1] 0.13590951 0.71223499 0.08027821

One-way Independent Groups ANOVA under Variance Inequality

- Welch's Independent Groups ANOVA
  - oneway.test (perf3 ~ group)
    - One-way analysis of means (not assuming equal variances)
    - data: perf3 and group

    - F = 3.0352, num df = 2.000, denom df = 57.694, p-value = 0.0558

Multiple Comparisons for Welch's Independent Groups ANOVA

- Multiplicity control with pairwise.t.test
  - pairwise.t.test(perf3, group, p.adjust.method = "hochberg", pool.sd=F)
  - pairwise.t.test(perf3, group, p.adjust.method = "by", pool.sd=F)
    - Pairwise comparisons using t tests with non-pooled SD
    - data: perf3 and group
    -
    -         1   2
    - 2 0.19   -
    - 3 1.00 0.15
    -
    - P value adjustment method: BY

One-way Independent Groups ANOVA under Nonnormality

- Kruskal-Wallis Nonparametric Test
  - kruskal.test (perf3 ~ group)
    - Kruskal-Wallis rank sum test

    - data: perf3 by group
    - Kruskal-Wallis chi-squared = 4.3842, df = 2, p-value = 0.1117
- Post hoc tests can be conducted with the 'wilcox.test' procedure and multiplicity control can be imposed with p.adjust

One-way Independent Groups ANOVA under Nonnormality and Variance Heterogeneity

- As in the two independent groups situation, we can use one of Rand Wilcox's functions (in this case t1way) for computing a Welch omnibus test on trimmed means
  - This test is much more reliable than a standard one-way ANOVA when the normality and variance homogeneity assumptions are violated

One-way Repeated Measures ANOVA

- Hypothesis #7: Is there a significant difference in perfectionism scores from pretest to one-month to posttest?
  - Problem: Simple methods for conducting repeated measures ANOVAs ignore the important sphericity assumption that is regularly violated with repeated measures data and inflates Type I error rates
  - Example:
    - mod5<- aov(perf ~ week + error (subject / week))
  - However, other functions are available in R that use adjusted df or multivariate solutions to solve the sphericity issue

One-way Repeated Measures ANOVA with the “car” package

– library(car)
– time<-c(1,2,3)
– time<-as.factor(time)
– idat<-data.frame(time)
– mod6<-lm(cbind(perf1,perf2,perf3)~1)
– aov1<-Anova(mod6, idata=idat, idesign=~time)
– summary(aov2)
  – Multivariate Tests: time
  – Df test stat approx F numDf denDf Pr(>F)
  – Pillai 1 0.290019 17.973521 2 88 2.85e-07 ***
  – Wilks 1 0.709981 17.973521 2 88 2.85e-07 ***
  – Roy 1 0.408489 17.973521 2 88 2.85e-07 ***
  –
  – Greenhouse-Geisser Correction for Departure from Sphericity
  – GG eps Pr(>F[GG])
  – time 0.68104 1.728e-07 ***

Follow-up Tests for a Repeated Measures ANOVA

- Follow-up tests can be conducted with two-sample paired t-tests and some sort of multiplicity control
  - p1<-t.test(perf1,perf2, paired=T)$p.value
  - p2<-t.test(perf1,perf3, paired=T)$p.value
  - p3<-t.test(perf2,perf3, paired=T)$p.value
  - p.adjust(c(p1,p2,p3),method="BH")

    – 1.336474e-07 2.000595e-06 1.972608e-02

Factorial Independent Groups ANOVA

- Hypothesis 8: Is there a significant relationship between posttest perfectionism scores and the predictors group and sex?
  - > sex<-as.factor(sex)
  - > anova(lm(perf3 ~ group + sex))
    - Analysis of Variance Table
-
    - Response: perf3
    -

	Df	Sum Sq	Mean Sq	F value	Pr(>F)	
group	2	3062	1531	2.7734	0.06804	.
sex	1	273	273	0.4937	0.48420	
Residuals	86	47477	552			

    -
    - Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Factorial Independent Groups ANOVA with an Interaction

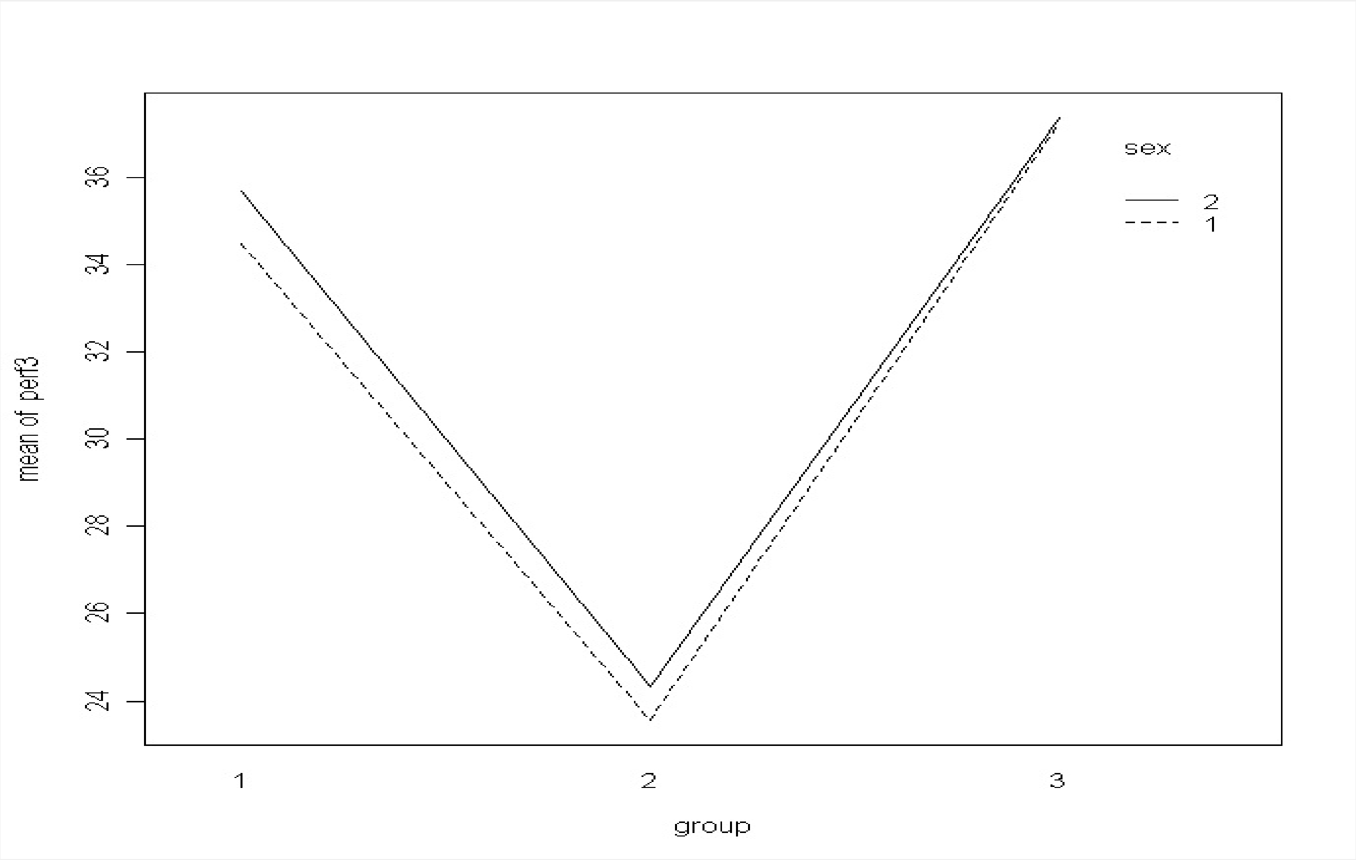
- > anova(lm(perf3 ~ group*sex))
  - Analysis of Variance Table
  - Response: perf3

	Df	Sum Sq	Mean Sq	F value	Pr(>F)	
group	2	3062	1531	2.6943	0.07343	.
sex	1	12	12	0.0210	0.88503	
group:sex	2	3	2	0.0028	0.99722	
Residuals	84	47735	568			

  - Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Factorial ANOVA: Plotting a Potential Interaction

- >interaction.plot(group, sex, perf3)



Mixed ANOVA

- Hypothesis 9: Are perfectionism scores affected by time, group, or the interaction of time and group
  - mod7<- lm(cbind(perf1,perf2,perf3)~group)
  - aov2<-Anova(mod7, idata=idat, idesign=~time)
  - summary(aov2)
    - Univariate Type II Repeated-Measures ANOVA Assuming Sphericity

	SS	numDf	Error SS	den Df	F	Pr(>F)	
group	4976	2	90585	87	2.3895	0.09766	.
time	3926	2	14054	174	24.3013	4.928e-10	
group:time	525	4	14054	174	1.6264	0.16969	