Initialize R

Initialize R by entering the following commands at the prompt. You must type the commands *exactly* as shown.

```r
options(contrasts=c("contr.sum","contr.poly"))  # set definition of contrasts
goats <- read.csv(file=url("http://psycserv.mcmaster.ca/bennett/psy710/datasets/goat.csv") )
closeAllConnections()
```

goats

The following text was taken from material posted at [www.statlab.uni-heidelberg.de/data/ancova/goats.story.html](http://www.statlab.uni-heidelberg.de/data/ancova/goats.story.html).

Experiments were carried out on six commercial goat farms to determine whether the standard worm drenching program was adequate. Forty goats were used in each experiment. Twenty of these, chosen completely at random, were drenched according to the standard program, while the remaining twenty were drenched more frequently. The goats were individually tagged, and weighed at the start and end of the year-long study. For the first farm in the study the resulting liveweight gains are given along with the initial liveweights. In each experiment the main interest was in the comparison of the liveweight gains between the two treatments.

The data from one of these experiments was taken from [www.statlab.uni-heidelberg.de/data/ancova/goats.data](http://www.statlab.uni-heidelberg.de/data/ancova/goats.data) and is stored in the data frame `goats`. Values of baseline (i.e., pre-treatment) weight and weight gain are stored in the variables `wt` and `gain`, respectively. Type of worm drenching – standard and intensive – is stored in the variable `treatment`.

linear regression

The following code uses linear regression to evaluate the linear relation between `gain` and `wt`:

```r
goats.lm.01 <- lm(gain~wt,data=goats)
summary(goats.lm.01)
```

```
# Call:  # lm(formula = gain ~ wt, data = goats)
#
# Residuals:  #   Min 1Q Median 3Q Max  # -3.607 -1.206 0.163 1.054 3.871
# # Coefficients:  # Estimate Std. Error t value Pr(>|t|)  # (Intercept) 14.39581 1.85047 7.780 2.22e-09 ***  # wt -0.35403 0.07906 -4.478 6.68e-05 ***
```
The `summary` function prints a regression table, which the regression coefficient, or $\beta$, for $\text{wt}$ as -0.35, which differs significantly from zero ($t(38) = -4.47, p < .0001$). This regression coefficient means that for every increase of 1 in $\text{wt}$, the value of $\text{gain}$ decreases (on average) by 0.35. The intercept is 14.39, and the $t$ test shows that it differs significantly from zero. Together, the intercept and regression coefficient define a straight line

$$\text{gain} = \text{intercept} - 0.354 \times \text{wt}$$

The overall fit of the regression model, as indexed by $R^2 = 0.34$, also is significant, $F(1,38) = 20.05, p < .0001$. $R^2$ is evaluated by noting the change in the goodness-of-fit that occurs when all of the parameters except the intercept are set to zero. In this case, there is only one parameter – the regression coefficient for $\text{wt}$ – and so you would think that the overall $F$ test should be related to the $t$ test for $\text{wt}$’s $\beta$ value... and you would be correct. In the case of a regression model that has only one regression coefficient, the overall $F$ equals $t^2$: $F = t^2 = -4.478^2 = 20.05$.

The following code uses `plot` to create a scatter plot of $\text{gain}$ vs. $\text{wt}$ and then uses the function `abline` to add the regression line defined by Equation 1. The resulting graph is shown in Figure 1.

```r
with(goats,plot(wt,gain,"p",xlab="wt",ylab="gain"))
abline(goats.lm.01)
```

**ANOVA & ANCOVA**

In this section we will use ANOVA and ANCOVA to evaluate the effect of treatment.

1. Conduct an ANOVA to evaluate the effect of treatment on gain. Calculate the strength of association between treatment and gain.

```r
with(goats,tapply(gain,treatment,mean)) # group means

## intensive  standard
## 6.85 5.55

goats.lm.02 <- lm(gain~treatment,data=goats)
anova(goats.lm.02) # anova table

## Analysis of Variance Table
##
## Response: gain
## Df Sum Sq Mean Sq F value Pr(>F)
## treatment 1 16.9 16.9000 4.1299 0.04916 *
## Residuals 38 155.5 4.0921
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

adj.r.squared <- summary(goats.lm.02)$adj.r.squared # short-cut
(omega.squared <- adj.r.squared) # approximation for 1-way design

## [1] 0.07429173
```
Figure 1: Plot of weight gain as a function of pre-treatment weight. The solid line shows the regression line.
**Answer:** The effect of treatment was significant ($F(1,38) = 4.13$, $p = 0.049$), so weight gain differed between the two groups.

2. Conduct an analysis of covariance (ANCOVA) that evaluates the effect of treatment after controlling for the linear association between wt and gain.

```r
options(contrasts=c("contr.sum","contr.poly") ) # make sure to use sum-to-zero effects
goats.lm.03 <- lm(gain~wt+treatment,data=goats) # anova model
anova(goats.lm.03) # anova table
```

## Analysis of Variance Table
### Response: gain
## Df Sum Sq Mean Sq F value Pr(>F)
## wt 1 59.548 59.548 22.7478 2.869e-05 ***
## treatment 1 15.995 15.995 6.1104 0.01816 *
## Residuals 37 96.857 2.618
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

**Answer:** An analysis of covariance revealed a significant effect of treatment ($F(1,37) = 6.11$, $p = 0.018$) after controlling for the linear association between weight gain and the covariate, wt ($F(1,37) = 22.7$, $p < 0.0001$).

3. Use the function `coef` to examine the coefficients, or parameters, of the ANCOVA model. What are the values of the parameters (i.e., the $\alpha$’s) for the two levels of treatment? Explain what the coefficients for the ANCOVA model mean.

```r
levels(goats$treatment)
## [1] "intensive" "standard"

coef(goats.lm.03)
## (Intercept) wt treatment1
## 14.3341789 -0.3513684 0.6324316

alpha.intensive <- 0.632;
alpha.standard <- 0 - 0.632;
```

**Answer:** The parameters define two straight lines, fit separately to the data from the intensive and standard treatment groups, that relate weight gain to pre-treatment weight. The two lines have the same slope, $m = -0.351$, which means that a 1 pound increase in pre-treatment weight was associated with a decrease of 0.35 pounds in the expected value of weight gain. The intercepts of the lines – i.e., the values where the lines intercept the y-axis – are 14.33 + 0.632 for the intensive group and 14.33 − 0.632 for the standard. To illustrate these points graphically, the following code uses the ANCOVA parameters to draw the lines fit to the standard and intensive groups. For comparison, the figure includes the regression line fit to all of the data.
Figure 2: Plot of weight gain as a function of pre-treatment weight for two treatment groups. The solid line shows the regression line fit to all of the data. The two dotted lines are the lines fit by the analysis of covariance; they are constrained to have the same slope.

4. Evaluate the ANCOVA’s homogeneity of slopes assumption.

```r
x.range <- range(goats$wt) # [min, max] for x-axis
y.range <- range(goats$gain) # [min, max] for y=axis
goats.intensive <- subset(goats, treatment=="intensive")
goats.standard <- subset(goats, treatment=="standard")
with(goats.intensive, plot(wt, gain, "p", pch=21, cex=2, xlim=x.range, ylim=y.range, xlab="wt", ylab="gain"))
with(goats.standard, points(wt, gain, pch=19)) # add points for standard group
abline(a=14.33+alpha.intensive, b=-0.351, lty=2) # line for intensive group
abline(a=14.33+alpha.standard, b=-0.351, lty=2) # line for standard group
abline(goats.lm.01) # overall regression line
# label lines:
text(x=28.5, y=5, "intensive")
text(x=27, y=4, "standard")
text(x=28, y=4.5, "all points")
# legend
legend(x=26, y=10.5, legend=c("intensive", "standard"), pch=c(21, 19))
```

```
goats.lm.04 <- lm(gain~wt+treatment+wt:treatment, data=goats)
anova(goats.lm.04)
## Analysis of Variance Table
##
## Response: gain
##  Df Sum Sq Mean Sq F value Pr(>F)
```
\[ \text{wt} \quad 159.548 \quad 59.548 \quad 22.2115 \quad 3.6e-05 \quad *** \\
\text{treatment} \quad 15.995 \quad 15.995 \quad 5.9663 \quad 0.01962 \quad * \\
\text{wt:treatment} \quad 0.342 \quad 0.342 \quad 0.1277 \quad 0.72296 \\
\text{Residuals} \quad 36 \quad 96.514 \quad 2.681 \\
\]---

\[ \text{Signif. codes: } 0 \ '***' \ 0.001 \ '**' \ 0.01 \ '*' \ 0.05 \ '.05' \ 0.1 \ ' ' \ 1 \]

\[
\text{coef(goats.lm.04)}
\]

\[ \text{(Intercept)} \quad \text{wt} \quad \text{treatment1} \quad \text{wt:treatment1} \\
14.36249702 \quad -0.35253363 \quad 0.01038434 \quad 0.02686777
\]

\[
\text{beta} \leftarrow -0.3525; \\
\text{alpha.intensive} \leftarrow 0.01038; \\
\text{alpha.standard} \leftarrow 0-0.01038; \\
\text{alpha.beta.intensive} \leftarrow 0.02686 \\
\text{beta} + \text{alpha.beta.intensive}
\]

\[ \# [1] -0.32564 \\
\text{beta} - \text{alpha.beta.intensive}
\]

\[ \# [1] -0.37936 \]

**Answer:** Like the ANCOVA model, the model that includes an interaction term fits two lines separately to each treatment group. Unlike the ANCOVA model, however, this more complex model allows the intercept and slope to vary across groups. Denote the coefficient for \( \text{wt} \) as \( \beta = -0.3525 \), and the interaction effects for the first and second treatment groups as \((\alpha\beta)_\text{intensive} = 0.02686\) and \((\alpha\beta)_\text{standard} = -0.02686\), respectively. Then the slopes of the lines fit to the first and second treatment groups are \(\beta + (\alpha\beta)_\text{intensive} = -0.379\) and \(\beta + (\alpha\beta)_\text{standard} = -0.325\). When the interaction effects are zero, the slopes for the two lines will be equal and the vertical separation between the two lines – i.e., the effect of treatment – will not depend on the level of the covariate. When the slopes differ, however, then the effects of treatment vary with the level of the covariate, and it may not be meaningful to talk about the effect of treatment. Fortunately, in the current case the \(\text{wt} \times \text{treatment} \) interaction was not significant \((F(1,37) = 0.127, p = 0.722)\), means that we have insufficient evidence to reject the null hypothesis that all of the interactions are zero. In addition to not being significant, the effect size for the interaction is small: when \( F < 1 \), it is common to set the association strength and effect size to zero. Therefore, we have good reason to accept the homogeneity assumption as being valid, and to examine effects of the treatment knowing that it does not depend strongly on the level of the covariate.

5. Calculate the **adjusted means** for the two treatment conditions.

```r
#install.packages(effects) # only need this if effects package isn't on your computer
library(effects)
effect(term="treatment",goats.lm.03)
```

\[
\#
\text{treatment effect}
\]
## treatment
## intensive  standard
## 6.832432  5.567568

6. Use a Tukey test to evaluate each pairwise difference between adjusted means. (N.B. There are only two
groups, so a Tukey test obviously is unnecessary in this case. Our purpose here is to learn how to do such
a test).

```r
#install.packages(multcomp) # only need this if multcomp not on your computer
library(multcomp)

## Loading required package: mutnorm
## Loading required package: survival
## Loading required package: TH.data

postHocs <- glht(goats.lm.03,linfct=mcp(treatment="Tukey") ) # pairwise Tukey tests
summary(postHocs) # print summary table with corrected p-values

##
## Simultaneous Tests for General Linear Hypotheses
##
## Multiple Comparisons of Means: Tukey Contrasts
##
## Fit: lm(formula = gain ~ wt + treatment, data = goats)
##
## Linear Hypotheses:
## Estimate Std. Error t value Pr(>|t|)
## standard - intensive == 0 -1.2649 0.5117 -2.472 0.0182 *
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## (Adjusted p values reported -- single-step method)

confint(postHocs) # confidence intervals

##
## Simultaneous Confidence Intervals
##
## Multiple Comparisons of Means: Tukey Contrasts
##
## Fit: lm(formula = gain ~ wt + treatment, data = goats)
##
## Quantile = 2.0262
## 95% family-wise confidence level
##
## Linear Hypotheses:
## Estimate lwr  upr
## standard - intensive == 0 -1.2649 -2.3017 -0.2281
```
7. Calculate the strength of association between treatment and gain using omega-squared and partial omega-squared. (See Section 9.3 of the course notes.)

```r
with(goats, tapply(gain, treatment, length))
```

```r
## intensive  standard
## 20 20
```

```r
N <- 20 + 20;
anova(goats.lm.03) # ancova table
```

```r
## Analysis of Variance Table
## Response: gain
## Df  Sum Sq Mean Sq F value  Pr(>F)
## wt   1 59.548  59.548 22.7478 2.869e-05 ***
## treatment 1 15.995  15.995  6.1104  0.01816 *
## Residuals 37 96.857  2.618
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
df.treat <- 1;
F.treat <- 6.11;
MS.treat <- 15.995;
MS.resids <- 2.618;
SS.total <- 59.548 + 15.995 + 96.857;
omega.squared <- (df.treat * (MS.treat - MS.resids)) / (SS.total + MS.resids)
partial.omega.squared <- (df.treat * (F.treat - 1)) / (df.treat * (F.treat - 1) + N)
```

```r
omega.squared
## [1] 0.07643214
partial.omega.squared
## [1] 0.1132787
```

(a) What is the difference between these two measures of association strength?

**Answer:** Omega-squared treats variation in the dependent variable that is associated with the covariate as part of error variance, whereas partial omega-squared removes variation associated with the covariate from error variance. **Question:** Do you see why partial omega-squared cannot be smaller than omega squared?

(b) How do these association strengths compare to the association strength estimated from the ANOVA in question 1?

**Answer:** Because omega-squared treats variation in the dependent variable that is associated with the covariate as part of error variance, it is identical to omega-squared estimated from the one-way ANOVA (which did not include a covariate term in the model). Partial omega-squared expresses association strength after controlling for variation associated with the covariate, and therefore is larger than omega-squared estimated from the one-way ANOVA.