Scientific Methods for Health Sciences: Applied Inference (HS851): Fall 2014

http://www.socr.umich.edu/people/dinov/2014/Fall/HS851

Homework 4 Solutions

Use the SOCR Neuroimaging dataset of visceral pain, irritable bowel syndrome, ulcerative colitis, and Crohn's disease.

Problem 1: Is there evidence of hippocampal volume asymmetry in the IBS cohort? Formulate an appropriate research hypothesis, identify a suitable statistical technique, apply the corresponding software tool, and report your findings.

H_0: The mean difference between right and left hippocampal volumes is equal to 0.
H_A: The mean difference between right and left hippocampal volumes is not 0.
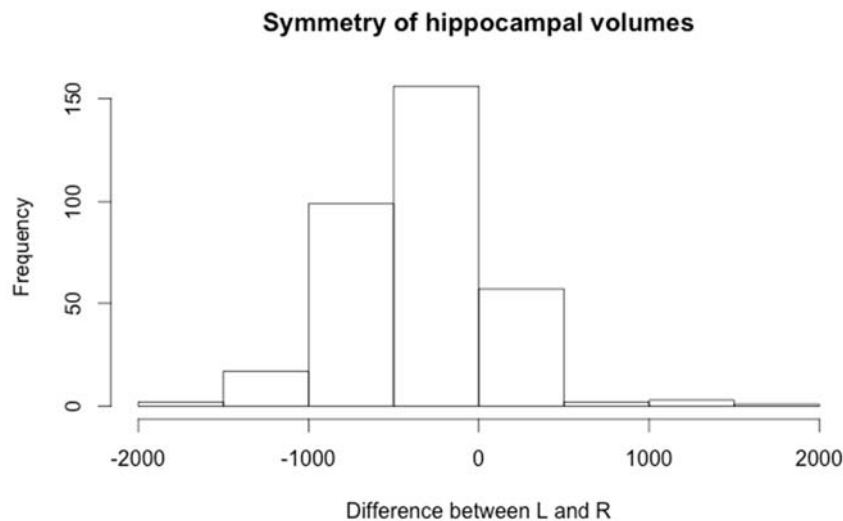
A histogram of the difference between the two data sets reveal that the right hippocampus is a bit larger on average. The differences appear relatively normally distributed. We apply a paired t-test to assess our hypothesis.

We find a t-value of -14.74 with 336 df, which results in a p-value of less than 2.2E-16. We therefore reject the null hypothesis and conclude that there is significant evidence to suggest that the right and left hippocampi have different volumes.

Problem 2: Assume the gender distributions of the participants in this study accurately represents the 4 cohorts (Normals, IBS, Ulcerative Colitis and Crohn's disease). Let's say we wish to compare the proportions of males in the Normals and IBS groups. Use the Methods of Moments to compute estimates (p_1 and p_2) of the probabilities that a randomly selected male is from the asymptomatic normal control group (p_1) and a randomly selected male is from the IBS control group (p_2). Once you find the point estimates (p_1 and p_2), perform a statistical test to assess the differences between the proportions of males in the Normals and IBS groups. Note that you need to invert the conditioning: p'_1=P(male|NC), and p'_2=P(male|IBS). What is the confidence interval of the difference $p'_1 - p'_2$?

	Male	Female	**Totals**
Normal	27	194	**221**
IBS	24	83	**107**
Totals	**51**	**277**	**328**

H_0 : The difference between the proportions of males in the normal and control groups equals 0.

H_A : The difference between the proportions of males in the normal and control groups is not equal to 0.

$\alpha = 0.05$

Assume a binomial model, $X \sim Bin(n,p)$, where n is the number of trials (i.e., total number of males) and p is the number of successes (i.e., asymptomatic for p1 and symptomatic for p2).

$$p_1 = P(NC|male) = \frac{27}{51} = 0.53$$
$$p_2 = P(IBS|male) = \frac{24}{51} = 0.47$$

E[X] = n*p
n*p1 = 27
n*p2 = 24
n=51
p1=27/51 = 0.53
p2=24/51 = 0.47

We ca use the inversion of conditioning to compute p'_1 and p'_2 (or we can compute it directly from the given data in the table).

$$p'_1 = P(male|NC) = P(NC|male)\frac{P(male)}{P(NC)} = \frac{27}{51}\frac{51/328}{221/328} = \frac{27}{51}\frac{51}{221} = 0.1221719,$$

$$p'_2 = P(male|IBS) = P(IBS|male)\frac{P(male)}{P(IBS)} = \frac{24}{51}\frac{51/328}{107/328} = \frac{24}{51}\frac{51}{107} = 0.2242991.$$

p'_1 = Proportion of males in normal group: 27/221 = 0.122
p'_2 = Proportion of males in IBS group: 24/107 = 0.224
Difference in proportions: (27/221)-(24/107) = -0.102
n1=248

n2=131

SE(p'1-p'2) = (sqrt(p'1*(1-p'1)/n1 + (p'2*(1-p'2)/n2))

Z-score: Zo = -0.102 / SE(p'1-p'2) = -2.22

P(Z<Zo) = 0.0132

Since 0.0132*2=0.0264<0.05

The p-value is less than α; therefore, we reject the null hypothesis and conclude that there is a significant difference between the proportion of males in the IBS group and that in the control group.

95% CI: [(p'1-p'2)-1.96*SE; (p'1-p'2)+1.96*SE)]= [-0.01207; -0.19218]

This interval does not include 0, so our confidence interval agrees with our test that the difference is significant.

Problem 3: Use the [SOCR Fligner-Killeen test](#) to investigate if the variances of the white matter volume (WMV) between the 3 cohorts (Normals, IBS, Ulcerative Colitis) are significantly different. Why is this important? Repeat this analysis looking for differences in the variability of the gray matter volume (GMV) in the controls group between genders. Elaborate on your findings.

It is important to know whether variances among groups are similar because many other statistical tests (e.g., ANOVA) rely on the assumption that variances among groups are equal.

WMV in 3 groups
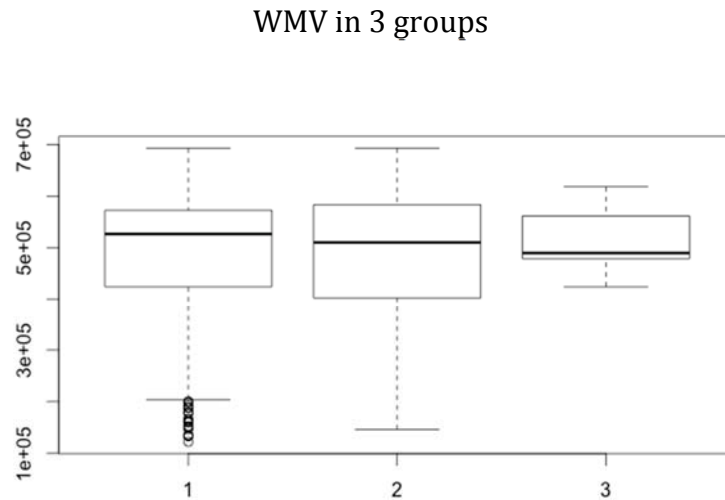
Group 1: Controls
Group 2: IBS
Group 3: Ulcerative colitis

R code:
fligner.test(list(wmv.1, wmv.2, wmv.3))

```
        Fligner-Killeen test of homogeneity of variances

data:  list(wmv.1, wmv.2, wmv.3)
Fligner-Killeen:med chi-squared = 3.0595, df = 2, p-value = 0.2166
```

The high p-value (0.2166) suggests that the variances in WMV among the three groups are not significantly different.

The parallel results from SOCR:

DATA   MAPPING   RESULT

Control	IBS	UC	C4	C5	C6	C7	C8	C9	C10	C11
581209	531989	423992								
580203	605453	506513								
459476	402158	478837								
477878	478254	617988								
539330	478210	616133								
563500	454652	478254								
213670	251391	489523								
657803	582628									
572350	524934									
510069	489523									
639823	573724									
633766	540942									
531363	571200									
292502	654870									
464631	549661									
529163	615259									
495913	620709									
542922	607804									
506513	588009									
588515	579160									
545699	504467									
533744	423992									
341898	583833									
556803	562218									
526386	544651									
515491	620158									
555772	543607									
467715	631145									
565614	488428									

DATA   MAPPING   RESULT

C4
C5
C6
C7
C8
C9
C10
C11
C12
C13
C14
C15
C16

SELECT AT LEAST TWO GROUPS:
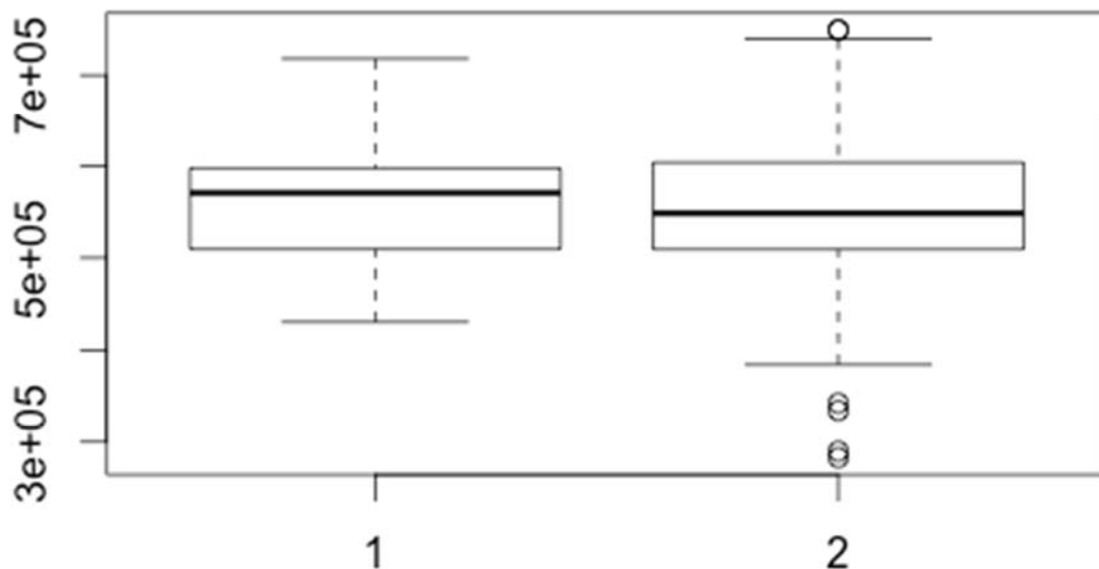ADD
REMOVE

Control
IBS
UC

Group Control
median = 526386.0

Group IBS
median = 510069.0

Group UC
median = 489523.0

Total Size = 335
Total Mean Score = .795
Total Variance = .354
Degrees of Freedom = 2
Chi-Square Statistic = 3.060
P-Value = .217

GMV among groups

> fligner.test(list(gmv.1, gmv.2))

        Fligner-Killeen test of homogeneity of variances

data: list(gmv.1, gmv.2)

Fligner-Killeen:med chi-squared = 0.277, df = 1, p-value = 0.5987

The high p-value (0.5987) suggests that the variances in GMV among the two groups are not significantly different.

The parallel analysis and results in SOCR:

| | | | CALCULATE | EXAMPLE 1 | EXAMPLE 2 | EXAMPLE 3 | EXAMPLE 4 | CLEAR | | |

| | | | | | | DATA | MAPPING | RESULT | |

males	females	C3	C4	C5	C6	C7	C8	C9	C10	C11
572172	673617									
585671	608842									
583833	566926									
516133	584741									
570818	630158									
512296	718007									
590993	605422									
583785	509249									
434763	508228									
430577	564193									
572234	582039									
453114	384022									
484690	536144									
528234	573103									
584741	476284									
718007	493260									
605422	506546									
564193	588193									
582039	569848									
474696	536474									
504467	566345									
548149	683785									
506546	548471									
569848	434763									
488280	664602									
560434	598462									
558783	518476									
537182	616013									
592302	560434									

Results:

Group C1
median = 553466.0

Group C2
median = 548471.0

Total Size = 337
Total Mean Score = .795
Total Variance = .354
Degrees of Freedom = 1
Chi-Square Statistic = .277
P-Value = .599

Problem 4: Complete the following table, where *patients* indicates all study participant except the asymptomatic controls.

Groups	Patients	Controls	Total
Female	89	194	283
Male	27	27	54
Total	221	116	337

Compute a point estimate and a 95% CI for the Odds Ratio of controls relative to gender. Next, apply the χ^2 association test to investigate for association between participants' gender and diagnosis.

Odds of being female given being a patient compared to the odds of being female given being a control.
OR: (89/27) / (194 / 27) = 0.46
log(OR) = -0.78
SE(log(OR)) = sqrt(1/89 + 1/27 + 1/194 + 1/27)
CI(log(OR)) = [log(OR) – 1.96 * SE, log(OR)+1.96*SE] = [-1.37, -0.19]
CI(OR) = exp (CI(log(OR)) = [0.25, 0.83]

The 95% confidence interval does not include 1, so we can conclude that the odds of being female is higher among controls than among patients.

Chi-squared test
> chisq.test(tab, correct=F)

      Pearson's Chi-squared test

data: tab
X-squared = 6.9136, df = 1, p-value = 0.008554

The low p-value of 0.009 suggests that the distribution of gender is not independent of disease status.

The parallel results from SOCR:

Results of Chi-Square Test for Independent or Homogeneity

Number of Rows = 2
Number of Columns = 2

	Patients	Controls	Row Total
Female	89.0 (97.412)	194.0 (185.588)	283
Male	27.0 (18.588)	27.0 (35.412)	54
Col Total	116	221	337

Degrees of Freedom = 1

Pearson Chi-Square Statistics = 6.914

P-Value = 0.00855415888759796

Appendix: R code

```
###################################
# HW 4
# Jennie Lavine
# 10.29.14
###################################
#The data, hw4.csv, was imported by copying the data from
#http://wiki.socr.umich.edu/index.php/SOCR_Data_April2011_NI_IBS_Pain,
#pasting it into an Excel spreadsheet,
#and saving as a comma-separated (i.e., .csv) file.

setwd('~/hw4_851')
dat<-read.csv('hw4.csv')


############################################
# Problem 1
############################################

mean(dat$L_hippocampus)
mean(dat$R_hippocampus)

par(mfrow=c(2,1))
hist(dat$L_hippocampus)
hist(dat$R_hippocampus)
par(mfrow=c(1,1))
hist(dat$L_hippocampus-dat$R_hippocampus, main='Symmetry of hippocampal volumes',
    xlab='Difference between L and R')

hippo_t.test <- t.test(x=dat$L_hippocampus, y=dat$R_hippocampus, paired=T)
hippo_t.test


#################################################
#Problem 2
#################################################
datsub<-subset(dat, dat$Group==1 | dat$Group==2)
datsub<-datsub[,c('Group','Sex')]
margin.table(table(datsub), 1)
margin.table(table(datsub), 2)
prop.table(table(datsub))

p1 <- 27/51
p2 <- 24/51

p1=27/221
p2=24/107
n1=27+221
```

```
n2=24+107

zscore = (p1-p2 / (sqrt( p1*(1-p1)/n1 + (p2*(1-p2)/n2)))
pnorm(zscore, 0, 1)

conf.int.95<- c(
  (p1-p2)-1.96*(sqrt( p1*(1-p1)/n1 + (p2*(1-p2)/n2))),
  (p1-p2)+1.96*(sqrt( p1*(1-p1)/n1 + (p2*(1-p2)/n2)))
  )




###############################################
#Problem 3
###############################################
wmv.1<-dat$WMV[dat$Group==1]
wmv.2<-dat$WMV[dat$Group==2]
wmv.3<-dat$WMV[dat$Group==3]

datsub<-subset(dat, dat$Group!=5)
datsub$Group<-factor(datsub$Group)
boxplot(WMV~as.factor(Group), data=datsub)

fligner.test(list(wmv.1, wmv.2, wmv.3))


gmv.1<-dat$GMV[dat$Group==1]
gmv.2<-dat$GMV[dat$Group==2]
gmv.3<-dat$GMV[dat$Group==3]

datsub<-subset(dat, dat$Group!=5)
datsub$Group<-factor(datsub$Group)
boxplot(GMV~as.factor(Group), data=datsub)

fligner.test(list(gmv.1, gmv.2, gmv.3))

#######################################
#Problem 4
#######################################
datsub<-dat[, c('Group','Sex')]
datsub$group.bin<-NA
datsub$group.bin[datsub$Group==1]<-1
datsub$group.bin[datsub$Group!=1]<-2
tab<-table(datsub[,2:3])
margin.table(tab,1)
margin.table(tab,2)
sum(tab)

OR <- (89*27)/(194*27)
```

```
log(OR)
se.log.or <- sqrt(sum(1/c(tab)))
ci.log.or<- c(log(OR)-1.96*se.log.or, log(OR)+1.96*se.log.or)
exp(ci.log.or)

chisq.test(tab, correct=F)
```