Hypothesis Testing 3

Day 13 (2/18/20)

4.2 Con't: The Behrens-Fisher problem (when $\sigma_1^2 \neq \sigma_2^2$)

Suppose that the following assumptions hold

1. $X_1, \ldots, X_{n_1} \sim N(\mu_1, \sigma_1^2)$ and $Y_1, \ldots, Y_{n_2} \sim N(\mu_2, \sigma_2^2)$, where $\sigma_1^2 \neq \sigma_2^2$.
2. The X sample and Y sample are independent

Welch (1947) showed that

$$\frac{(\overline{X} - \overline{Y}) - (\mu_1 - \mu_2)}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}}$$

has an approximate t-distribution with degrees of freedom close to

$$v^* = \frac{\left(\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}\right)^2}{\frac{S_1^4}{n_1^2(n_1-1)} + \frac{S_2^4}{n_2^2(n_2-1)}}$$

Then using the pivot method again,

$$\begin{aligned} .95 &= P\left[-t_{.025,v^*} \leq \frac{(\overline{X} - \overline{Y}) - (\mu_1 - \mu_2)}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}} \leq t_{.025,v^*}\right] \\ &\vdots \\ &= P\left[(\overline{X} - \overline{Y}) - t_{.025,v^*}\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}} \leq \mu_1 - \mu_2 \leq (\overline{X} - \overline{Y}) + t_{.025,v^*}\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}\right] \end{aligned}$$

so that a 95% confidence interval for $\mu_1 - \mu_2$ is given by

$$(\overline{X} - \overline{Y}) \pm t_{.025,v^*}\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}$$

This is called the Welch-Satterthwaite (or simply Welch) confidence interval for the difference between two means. This is the default method for the *t.test()* function in R.

Example(Exercise 4.6.6: AZT data)

```
> load(url('http://www.stat.wmich.edu/~mckean/hmchomepage/Data/aztdoses.rda'))
> head(aztdoses)
  azt dose
1 284  300
2 279  300
3 289  300
```

```
4 292  300
5 287  300
6 295  300
> attach(aztdoses)
> x1<-azt[dose==300]
> x2<-azt[dose==600]
> x1
 [1] 284 279 289 292 287 295 285 279 306 298
> x2
 [1] 298 307 297 279 291 335 299 300 306 291

> t.test(x1,x2)

Welch Two Sample t-test

data:  x1 and x2
t = -2.034, df = 14.509, p-value = 0.06065
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -22.3557409   0.5557409
sample estimates:
mean of x mean of y
    289.4     300.3

> t.test(x1,x2,var.equal=T)

Two Sample t-test

data:  x1 and x2
t = -2.034, df = 18, p-value = 0.05696
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -22.1584072   0.3584072
sample estimates:
mean of x mean of y
    289.4     300.3
```

In general, the Welch-Satterthwaite $(1-\alpha)100\%$ confidence interval for $\mu_1 - \mu_2$ is

$$(\overline{X} - \overline{Y}) \pm t_{\alpha/2,v^*}\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}$$

Comments:

1. Welch-t requires normality, but relaxes the equal variance assumption required by pooled-t

2. If $\sigma_1^2 = \sigma_2^2$, Welch-t and pooled-t have approximately the same performance. If $\sigma_1^2 \neq \sigma_2^2$, Welch-t is better because pooled-t confidence interval may not have the desired coverage probability

3. Simulations suggest that conducting the Welch-t all the time is better than the two-stage test

   - Conduct a test for equal variance
   - Conduct pooled-t or Welch-t depending on outcome of test for equal variance

   This is because the test for equal variance is not sensitive enough, and the pooled-t ends up getting used even when variances are not equal

4. When normality is violated and sample sizes are small, then use nonparametric methods like the rank sum test (from Stat 5660)

5 Dependent Means and Proportions

5.1 Paired data

Let $(X_1, Y_1), \dots (X_n, Y_n)$ be a random sample of paired observations, where $(X_1, \dots, X_n)$ is a random sample from a distribution $f_1(\cdot)$ with mean μ_1 and $(Y_1, \dots, Y_n)$ is a random sample from a distribution $f_2(\cdot)$ with mean μ_2. Then the differences $(D_1, \dots, D_n) = Y_1 - X_1, \dots, Y_n - X_n$ constitute a random sample from a distribution $g(\cdot)$ with mean $\mu_2 - \mu_1$ (call this μ_d).

To test $H_0 : \mu_1 = \mu_2$ vs $H_1 : \mu_i \neq \mu_2$, we instead test

$$H_0 : \mu_d = 0 \text{ versus } H_1 : \mu_d \neq 0$$

which is a one-sample problem on the differences $D_1, \dots, D_n$. Consequently, we can construct a 95% confidence interval for μ_d as follows:

$$\overline{D} \pm t_{.025,n-1} \frac{S_d}{\sqrt{n}}$$

A test with level of significance $\alpha = .05$ will reject $H_0 : \mu_d = 0$ if

$$\left| \frac{\overline{D}}{S_d/\sqrt{n}} \right| \geq t_{.025,n-1}$$

5.2 Paired binary data

Example: BMR handout

		Week 2	
		N	Ab
Week 0	N	5	4
	Ab	1	0

In general, we express the data as

		Post	
		N	Ab
Pre	N	a	b
	Ab	c	d

McNemar's test: To test

$$H_0 : p_1 = p_2 \text{ versus } H_1 : p_1 \neq p_2$$

where p_1 and p_2 are the marginal probabilities (of "Normal", say), reject H_0 if

$$M = \frac{(b-c)^2}{b+c} > \chi^2_{.05,1}$$

Example: (BMR handout con't.) Since $M = \frac{(4-1)^2}{4+1} = \frac{9}{5} = 1.8$ which is not greater than $\chi^2_{.05,1} = 3.84$, then we do not reject the null hypothesis. The percentage of normal (or abnormal) is not significantly different between Week 0 and Week 2.

5.3 Dependent proportions

When two proportions $\hat{p}_1$ and $\hat{p}_2$ are two categories of a multinomial, then the two proportions are not independent. In fact, the sum $\hat{p}_1 + \hat{p}_2$ cannot exceed 1.0, so when one exceeds .5, then the other cannot. In this case, the estimator of $p_1 - p_2 = \hat{p}_1 - \hat{p}_2$ is the same as before, but the variance formula is different.

$$\text{Var}(\hat{p}_1 - \hat{p}_2) = \frac{\hat{p}_1(1-\hat{p}_1)}{n} + \frac{\hat{p}_2(1-\hat{p}_2)}{n} + \frac{2\hat{p}_1\hat{p}_2}{n}$$

Example: Suppose that $n = 50$ people were asked whether they are optimistic about the economy. The data is shown below. Are there significant more 'yes' than 'no'?

Yes	No	Not sure
22	15	13

Solution:

$$\hat{p}_1 - \hat{p}_2 = \frac{22}{50} - \frac{15}{50} = .44 - .30 = .14$$

$$\text{Var}(\hat{p}_1 - \hat{p}_2) = \frac{(.44)(.56)}{50} + \frac{(.30)(.70)}{50} + \frac{(.44)(.30)}{50} = .0118$$

The standard error of $\hat{p}_1 - \hat{p}_2 = .14$ is

$$\text{SE}(\hat{p}_1 - \hat{p}_2) = \sqrt{.0118} = .11$$

The 95% confidence interval of the difference is $.14 \pm 1.96(.11)$ or

$$(-.08, .25)$$

Appendix: R simulation of Welch versus pooled-t

```
n1<-35                      # Sample size for x
n2<-25                      # Sample size for y
sigma1<-50                  # Population sd for x
sigma2<-50                  # Population sd for y
mu1<-40                     # Population mean for x
mu2<-0                      # Population mean for y
nsim<-10000                 # Number of trials
pval1<-numeric(nsim)        # Storage for p-value of pooled-t
pval2<-numeric(nsim)        # Storage for p-value of Welch-t
for(i in 1:nsim){
xsim<-rnorm(n1,mu1,sigma1)  # Generate x-data
ysim<-rnorm(n2,mu2,sigma2)  # Generate y-data
pval1[i]<-t.test(xsim,ysim,alternative="two.sided",var.equal=TRUE)$p.value
pval2[i]<-t.test(xsim,ysim,alternative="two.sided",var.equal=FALSE)$p.value
}
cbind(mean(pval1<.05),mean(pval2<.05))

# Write a function
welch_sim<-function(nsim=10000, n1=30, n2=30, mu1=0, mu2=0, sigma1=1,sigma2=1){
pval1<-numeric(nsim)        # Storage for p-value of pooled-t
pval2<-numeric(nsim)        # Storage for p-value of Welch-t
for(i in 1:nsim){
xsim<-rnorm(n1,mu1,sigma1)  # Generate x-data
ysim<-rnorm(n2,mu2,sigma2)  # Generate y-data
pval1[i]<-t.test(xsim,ysim,alternative="two.sided",var.equal=TRUE)$p.value
pval2[i]<-t.test(xsim,ysim,alternative="two.sided",var.equal=FALSE)$p.value
}                           # End of for() loop
return(c(mean(pval1<.05),mean(pval2<.05)))
}                           # End of function()

# Call the function
welch_sim(10000,35,25,0,0,50,50)
welch_sim(10000,35,25,10,0,50,50)
welch_sim(10000,35,25,20,0,50,50)
welch_sim(10000,35,25,30,0,50,50)
welch_sim(10000,35,25,40,0,50,50)
# Use unequal variance
welch_sim(10000,35,25,mu1=0,0,20,50)
welch_sim(10000,35,25,mu1=0,0,50,20)

# Use for() loop to automate
muvec<-c(-40,-35,-30,-25,-20,-15,-10,-5,0,5,10,15,20,25,30,35,40)
outmat<-matrix(rep(0,2*length(muvec)), ncol=2)
for(j in 1:length(muvec)){
```

```
  outmat[j,]<-welch_sim(10000,35,25,muvec[j],0,20,50)
}
outmat
colnames(outmat)<-c("pooled","welch")
cbind(muvec,outmat)
power1<-outmat[,1]
power2<-outmat[,2]
plot(power1~muvec,ylim=c(0,1),type="l",xlab="mu1-mu2",ylab="Power",
     main="Power curves")
lines(power2~muvec,col="red",type="l",lty=2)
legend("topleft",legend=c("pooled","welch"),lty=c(1,2))
abline(h=.05,col="blue",lty=6)
```