R code for inference (confidence interval, hypothesis testing) about two proportion.

Hypothesis testing and P-values:
Example: Out of 29 males sampled, 9 think their sexual partner is responsible to ask about safer sex. Out of the 31 females sampled, 4 think their sexual partner is responsible to ask about safer sex. Are the proportions for males and females who think their partner is responsible for asking about safer sex <u>significantly different</u>?

We test the hypotheses $H_0: \pi_m - \pi_f = 0$ versus the alternative hypothesis $H_A: \pi_m - \pi_f \neq 0$

```
> prop.test(x=c(9,4),n=c(29,31),correct=FALSE)

        2-sample test for equality of proportions without continuity
        correction

data:  c(9, 4) out of c(29, 31)
X-squared = 2.9022, df = 1, p-value = 0.08846
alternative hypothesis: two.sided
95 percent confidence interval:
 -0.02430281  0.38692795
sample estimates:
   prop 1    prop 2
0.3103448 0.1290323
```

Conclusion: Since the P-value (0.0885) is large (larger than our alpha level of 5%), we fail to reject the null hypothesis. We have insufficient evidence to conclude the two proportions for males and females are significantly different.

In hindsight, we can ask how powerful our test is for given values of the true proportions of males and females. For this, I have written a basic power function that computes the power (see next page). For instance, suppose $\pi_m = 30\%$ and $\pi_f = 10\%$. How powerful is our test to detect this difference of 20 percent points with our samples:

```
> power(true.pi1=0.3, true.pi2=0.1, y=c(9,4), n=c(29,31))
[1] 0.4660046
```
Let's find and plot the power when π_m varies between 0.2 and 0.5:

```
> power.seg = power(true.pi1=seq(0.2,0.5,0.01), true.pi2=0.1,
y=c(9,4), n=c(29,31))
> power.seg
 [1] 0.1186957 0.1444755 0.1729777 0.2040027 0.2372961 0.2725574
 [7] 0.3094514 0.3476181 0.3866835 0.4262699 0.4660046 0.5055289
[13] 0.5445046 0.5826200 0.6195947 0.6551828 0.6891755 0.7214016
[19] 0.7517279 0.7800586 0.8063334 0.8305257 0.8526392 0.8727056
[25] 0.8907808 0.9069414 0.9212809 0.9339066 0.9449357 0.9544923
[31] 0.9627041
> plot(power.seg~seq(0.2,0.5,0.01), type="l", main="Power for
increasing values of true.pi1", xlab="true.pi1", ylab="Power")
```

Power for increasing values of true.pi1
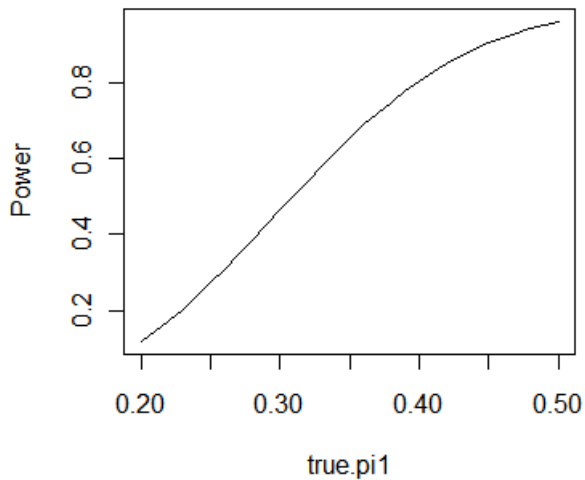
Here is the R code for the power function:

```
#######################################################
# This gives the power of the score test              #
# for testing H0: true.pi1 - true.pi2 = 0             #
# versus HA: true.pi1 - true.pi2 unequal (or > or <) 0 #
#######################################################

power <- function(true.pi1, true.pi2, y1, n1, y2, n2, alpha=0.05, alternative="not
equal") {
  z.q = switch(alternative,
        "less" = qnorm(alpha),
        "greater" = qnorm(1-alpha),
        qnorm(1-alpha/2)
        )
  p0 = (y1+y2)/(n1+n2)
  luis = (z.q*sqrt(p0*(1-p0)*(1/n1 +1/n2))  - (true.pi1-true.pi2)) /
sqrt(true.pi1*(1-true.pi1)/n1 + true.pi2*(1-true.pi2)/n2)
  if (alternative=="not equal") luis1 = (-z.q*sqrt(p0*(1-p0)*(1/n1 +1/n2))  -
(true.pi1-true.pi2)) / sqrt(true.pi1*(1-true.pi1)/n1 + true.pi2*(1-true.pi2)/n2)
  pow = switch(alternative,
        "less" = pnorm(luis),
        "greater" = 1-pnorm(luis),
        pnorm(luis1) + (1-pnorm(luis))
        )
  return(pow)
}
```

Confidence Intervals:

A confidence interval is much more informative than a simple P-value and the decision to reject or not reject. It tells about the size of the difference, and so it's easy to see if the difference is practically relevant or not (even if it is statistically significant).

The R function prop.test() does not give the score interval I have discussed in class. However, it is always best to use the score interval. I have written R code that computes this interval. Don't worry if you don't understand what the code is doing, it is based on a paper cited in the code. The principle is as we discussed in class, one simply inverts the score test and all values that are not rejected by the two-sample score test are part of the confidence interval. The code is also on our website.

```
score.sig <- function(delta,p,n,comb,z2) {
#based on relationship/identity in Andres et al. (2011) Stat. Meth. Med. Research
#if function value <0 than score statistic is not significant at 1-conflev level
  C= z2/(sum(comb*p) - delta)
  b=1-2*p
  R=sqrt(n^2+2*n*comb*b*C+comb^2*C^2)
  sum(n)+(sum(comb)-2*delta)*C-sum(R)
}

score.int <-function(y,n,conflev=0.95, alternative="two-sided", comb=c(1,-1)) {
  delta.max=sum(comb[comb>=0])
  delta.min=sum(comb[comb<0])
  p=y/n
  point.est=sum(comb*p)
  eps=10^(-5)
  if (alternative=="two-sided") {
      z=qnorm(1-(1-conflev)/2)
      if(point.est==delta.min) delta.lb=delta.min   #to handle y=0 or y=n
      else delta.lb=uniroot(score.sig,interval=c(delta.min,point.est-
eps),p=p,n=n,comb=comb,z2=z^2)$root
      if(point.est==delta.max) delta.ub=delta.max   #to handle y=0 or y=n
      else
delta.ub=uniroot(score.sig,interval=c(point.est+eps,delta.max),p=p,n=n,comb=comb,z2
=z^2)$root
      return(c(delta.lb,delta.ub))
      }
  else {
      z=qnorm(1-(1-conflev))
      if (alternative=="lower") {
        if(point.est==delta.min) delta.lb=delta.min   #to handle y=0 or y=n
        else delta.lb=uniroot(score.sig,interval=c(delta.min,point.est-
eps),p=p,n=n,comb=comb,z2=z^2)$root
      return(delta.lb)
      }
      else {
        if(point.est==delta.max) delta.ub=delta.max   #to handle y=0 or y=n
        else
delta.ub=uniroot(score.sig,interval=c(point.est+eps,delta.max),p=p,n=n,comb=comb,z2
=z^2)$root
        return(delta.ub)
        }
      }
}
```

Let's find the confidence interval for the difference of proportions for the example above:

> score.int(y=c(9,4), n=c(29,31))
[1] -0.02850764   0.38736729

Conclusion: We are 95% confident that the proportion of males asking their partner for safer sex is at least 2.8 percent points smaller and at most 38.7 percent points larger that the proportion for females. Since zero is included in the interval, the two proportions are not significantly different.

Note: To get confidence intervals with a different confidence coefficient, use `conflev=0.90` or `conflev=0.99` when calling the function:

```
> score.int( y=c(9,4), n=c(29,31), conflev=0.9)
[1] 0.006494263  0.354690884
```

So, with 90% confidence, the two proportions are significantly different and the proportion for males is by at least 0.6 percent points (almost negligible) and by at most 35.5 percent points larger than the female proportion.

To get one-sided confidence intervals, use the option alternative="lower" (for a lower bound) or alternative="upper":

```
> score.int( y=c(9,4), n=c(29,31), alternative="lower",
conflev=0.90)
[1] 0.04600412
```

If you don't have the above code available, you can use the built in prop.test, which often gives similar results when n1 and n2 are larger than roughly 20 and the sample proportions are not too close to 0 or 1:

```
> prop.test(x=c(9,4), n=c(29,31), correct=FALSE)

        2-sample test for equality of proportions without continuity
correction

data:  c(9, 4) out of c(29, 31)
X-squared = 2.9022, df = 1, p-value = 0.08846
alternative hypothesis: two.sided
95 percent confidence interval:
 -0.02430281  0.38692795
sample estimates:
   prop 1    prop 2
0.3103448 0.1290323
```

Compare this to the 95%, two-sided score interval from above: [1] -0.02850764 0.38736729