

## Statistics 406 Problem Set 6

Due in lab, Tuesday October 31st

1. The two-sample t-statistic given in the course notes is recommended when the two populations being compared have approximately equal variance. In this problem you will use simulation to see how sensitive the test is to this assumption.

Use simulation to get the actual level of the two-sample t-test (the one given in the notes that uses the pooled variance). Your test should be two-sided, with a nominal level of 0.05. Consider three situations for the sample sizes: 10,10; 10,20; and 10,100. Allow the ratio between the two population variances to take on each value in the following list: 1/5, 1/3, 1/2, 1, 2, 3, and 5.

In addition to the simulation code and numerical results, briefly discuss the circumstances in which the nominal and actual levels of the test agree and do not agree.

### Solution:

```
## Sample sizes for the two groups.
m <- 100
n <- 10

## Storage for the result.
P <- NULL

## Loop through the variance ratios.
for (f in c(1/5, 1/3, 1/2, 1, 2, 3, 5))
{
  ## Simulate 10000 data sets from each population.
  X <- array(rnorm(10000*m), c(m,10000))
  Y <- sqrt(f)*array(rnorm(10000*n), c(n,10000))

  ## Means and variances of each simulated dataset.
  XM <- apply(X, 2, mean)
  YM <- apply(Y, 2, mean)
  VX <- apply(X, 2, var)
  VY <- apply(Y, 2, var)

  ## The pooled variance estimate.
  Sp2 <- ((m-1)*VX + (n-1)*VY) / (m+n-2)

  ## The test statistic.
```

```

T <- abs(XM - YM) / sqrt(Sp2*(m+n)/(m*n))

## The p-value.
pv <- 2*(1 - pt(T, 2*n-2))

## Save the result.
P <- c(P, mean(pv < 0.05))
}

```

This is what I get:

```

100/10:
0.0001 0.0012 0.0066 0.0369 0.1188 0.1788 0.2545

20/10:
0.0140 0.0195 0.0263 0.0433 0.0708 0.0876 0.1152

10/10:
0.0579 0.0549 0.0512 0.0505 0.0524 0.0514 0.0522

```

The actual levels are very close to the nominal level (0.05) if the sample sizes are equal, regardless of the variances, and also if the variances are equal, regardless of the sample sizes. If both the sample sizes and variances are highly unequal in the two populations being compared, the actual level can be quite far from 0.05. For example, if population 1 has sample size 100 and variance 1, and population 2 has sample size 10 and variance 5, then the level is 0.25. This means that 25% of the time the null hypothesis of equal means is rejected.

- Suppose that paired data are generated as follows: (i) generate  $Z_i$  as a normal value with mean zero and variance  $\sigma^2$ , (ii) generate  $X_i = Z_i + a_i$  and  $Y_i = c + Z_i + b_i$ , where  $a_i$  and  $b_i$  are independent of each other and follow a standard normal distribution (and are independent of  $Z_i$ ), and  $c$  is a constant. The observed paired data are  $X_i, Y_i$ . The values of  $Z_i, a_i, b_i$  and  $c$  are not directly observable.

The following two paragraphs should help to explain the motivation for this problem, but may not be directly helpful in solving it.

In real terms, the values of  $X_i$  and  $Y_i$  represent measurements made on an individual before and after a treatment. The value of  $Z_i$  represents properties unique to individual  $i$  that are not affected by the treatment, but that are permanent (therefore affect  $X_i$  and  $Y_i$  equally).

The values of  $a_i$  and  $b_i$  represent properties of individual  $i$  that are transient, and therefore affect only one of the measurements. The treatment effect is captured by the constant  $c$ .

For example, suppose we are looking at a treatment for high blood pressure. Then  $X_i$  and  $Y_i$  would be blood pressure readings made before and after the treatment period,  $Z_i$  would reflect things like age, weight, and diet (which affect blood pressure, but aren't directly affected by the treatment), and  $a_i$  and  $b_i$  would represent things like the mood of the subject on the day of the reading, or whether the subject had recently eaten or ingested any caffeine (these things would temporarily alter the subject's blood pressure).

(a) What is  $E(Y_i - X_i)$ ?

**Solution:**  $E(Y_i - X_i) = c$ , which is the treatment effect.

(b) What is  $\text{cov}(X_i, Y_i)$

**Solution:**

$$\begin{aligned}\text{cov}(X_i, Y_i) &= \text{cov}(Z_i + a_i, Z_i + b_i + c) \\ &= \sigma^2\end{aligned}$$

(c) What is  $\text{cor}(X_i, Y_i)$ ?

**Solution:**

$$\text{var}(X_i) = \text{var}(Y_i) = 1 + \sigma^2,$$

therefore

$$\text{cor}(X_i, Y_i) = \sigma^2 / (1 + \sigma^2).$$

(d) Use simulation to support your results from parts 2 and 3.

**Solution:** When  $\sigma^2 = 0, 1, 100$ , the covariance is 0, 1, 100, and the correlation is 0, 1/2, 100/101, according to the results from above. This is consistent with the results of the following simulation.

```
## The treatment effect.
c <- 1

## Storage for the results.
V <- NULL
```

```

## Loop over several possible variance values.
for (sigma2 in c(0, 1, 100))
{
  ## The unobserved random value that is individual-specific and is
  ## not effected by the treatment.
  Z <- sqrt(sigma2) * rnorm(10000)

  ## The observed data.
  X <- Z + rnorm(10000)
  Y <- Z + rnorm(10000) + c

  ## Store the covariance and correlation.
  V <- rbind(V, c(cov(X,Y), cor(X,Y)))
}

```

- (e) Use simulation to estimate the power of the unpaired one-sided two-sample t-test for detecting a treatment effect  $c = 0.1$  based on sample sizes 10, 20, and 30.

**Solution:**

```

## The treatment effect size.
c <- 0.5

## Storage for the results.
Power <- NULL

## Loop over the sample sizes.
for (n in c(10, 20, 30))
{
  ## Loop over the variances for Z.
  for (sigma2 in c(0, 1, 3))
  {
    ## Generate 1000 sets of paired X and Y data of sample size n.
    Z <- array(sqrt(sigma2)*rnorm(n*1000), c(n,1000))
    a <- array(rnorm(n*1000), c(n,1000))
    b <- array(rnorm(n*1000), c(n,1000))
    X <- Z + a
    Y <- Z + b + c

    ## Get the numerators of the 1000 t-test statistics.
    MX <- apply(X, 2, mean)
    MY <- apply(Y, 2, mean)

```

```

D <- MY - MX

## Get the pooled variance estimates for the 1000 data sets.
VX <- apply(X, 2, var)
VY <- apply(Y, 2, var)
Sp2 <- ((n-1)*VX + (n-1)*VY) / (2*n-2)

## The 1000 t-test statistics.
T <- D / sqrt(Sp2*2/n)

## The 1000 p-values.
pv <- 1 - pt(T, 2*n-2)

## Save the sample size, variance, and power. The power is estimated
## as the proportion of the 1000 p-values that are less than 0.05.
Power <- rbind(Power, c(n, sigma2, mean(pv < 0.05)))
}
}

```

I get the following:

```

      [,1] [,2] [,3]
[1,]   10    0 0.302
[2,]   10    1 0.118
[3,]   10    3 0.026
[4,]   20    0 0.445
[5,]   20    1 0.211
[6,]   20    3 0.059
[7,]   30    0 0.620
[8,]   30    1 0.344
[9,]   30    3 0.099

```

This shows that for a given sample size, the unpaired statistic has decreasing power as the variance of  $Z$  increases. Thus the greater the tendency of before treatment and after treatment measures to be similar within a subject, the poorer the unpaired statistic performs in terms of power. If we fix the variance of  $Z$  and look across the different sample sizes, power increases with sample size as expected.

(f) Now repeat part (d) using a paired t-test.

**Solution:**

```
## The treatment effect size.
```

```

c <- 0.5

## Storage for the result.
Power <- NULL

## Loop over the sample size.
for (n in c(10, 20, 30))
{
  ## Loop over the Z variance.
  for (sigma2 in c(0, 1, 3))
  {
    ## Simulate 1000 paired data sets of size n.
    Z <- array(sqrt(sigma2)*rnorm(n*1000), c(n,1000))
    a <- array(rnorm(n*1000), c(n,1000))
    b <- array(rnorm(n*1000), c(n,1000))
    X <- Z + a
    Y <- Z + b + c

    ## Calculate the paired test statistic and p-value.
    D <- Y - X
    M <- apply(D, 2, mean)
    SD <- apply(D, 2, sd)
    T <- sqrt(n)*M / SD
    pv <- 1 - pt(T, n-1)

    ## Store the sample size, variance of Z, and the power. The power
    ## is estimated as the proportion of p-values below 0.05.
    Power <- rbind(Power, c(n, sigma2, mean(pv < 0.05)))
  }
}

```

I get the following:

```

      [,1] [,2] [,3]
[1,]  10   0 0.265
[2,]  10   1 0.264
[3,]  10   3 0.278
[4,]  20   0 0.457
[5,]  20   1 0.468
[6,]  20   3 0.461
[7,]  30   0 0.614
[8,]  30   1 0.590
[9,]  30   3 0.594

```

When  $\text{var}(Z) = 0$ , the powers are very close to those for the unpaired test. However unlike the unpaired test, the power does not declines as  $\text{var}(Z)$  increases.

- (g) Both the paired and unpaired test statistics have  $\bar{Y} - \bar{X}$  in their numerator. What is the expected value of this quantity?

**Solution:** The expected value of the numerator of the test statistic is the treatment effect:

$$E(\bar{X} - \bar{Y}) = c$$

- (h) What is the approximate expected value of the pooled variance estimate  $S_p^2$  (to get the approximate expected value, assume  $\bar{X} = EX$  and  $\bar{Y} = EY$ )?

**Solution:** For paired data,

$$S_p^2 \approx \left( \sum_i (X_i - EX)^2 + \sum_i (Y_i - EY)^2 \right) / (2n - 2)$$

and for any random variable  $U$ ,  $E(U - EU)^2 = \text{var}(U)$ . Therefore

$$\begin{aligned} S_p^2 &\approx (n\text{var}(X) + n\text{var}(Y)) / (2n - 2) \\ &= 2n(1 + \sigma^2) / (2n - 2) \\ &\approx 1 + \sigma^2 \end{aligned}$$

Therefore the expected value of  $S_p^2$  will be close to  $\sigma^2 + 1$ .

- (i) Approximately what will be the value of  $\hat{\sigma}^2$  that arises in the denominator of the paired test?

**Solution:** For the paired test,  $\hat{\sigma}^2$  is the sample variance of

$$X_i - Y_i = c + a_i + b_i,$$

which is 2.

- (j) Explain the difference in power between the paired and unpaired tests in terms of your answers to parts f, g, and h.

**Solution:** Both statistics will have a numerator around  $c$ . However the denominators are different. The denominator of the paired test statistic will be approximately  $\sqrt{2/n}$ , whereas the denominator of the unpaired test statistic will be approximately

$\sqrt{2(1 + \sigma^2)/n}$ . Thus when  $\sigma^2 = 0$  the two tests perform very similarly. However as  $\sigma^2$  gets larger than zero, the unpaired test statistics grow smaller (the denominator grows larger), and therefore it is less likely to reject. The paired statistic is not affected by the value of  $\sigma^2$ .