

Statistics 641, Fall 2011
Homework #2
Answers

1. Suppose that we have 8 subjects in each of two groups. We observe the following responses:
Control: 0.2 0.8 1.9 2.2 2.6 3.9 8.2 21.8
Experimental: 2.8 5.1 7.1 7.7 12.3 18.8 27.1 39.7

- (a) Calculate by hand the Wilcoxon rank sum test statistic for the comparison of the two groups.

Use “C” and “E” to denote the two groups. The observations are ordered as follows:

	0.2	0.8	1.9	2.2	2.6	2.8	3.9	5.1	7.1	7.7	8.2	12.3	18.8	21.8	27.1	39.7
group:	C	C	C	C	C	E	C	E	E	E	C	E	E	C	E	E
rank:	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16

The sum of the ranks in group “E” is $T = 6 + 8 + 9 + 10 + 12 + 13 + 15 + 16 = 89$. The expected rank in group E is $8 \times 17/2 = 68$, so the $T - E[T] = 89 - 68 = 21$.

Because there are no ties, the variance is $8 \times 8 \times 17/12 = 90.667$. The chi-square test statistic is

$$\frac{21^2}{90.667} = 4.86$$

This statistic has (asymptotically) a chi-square distribution with 1 df. The p -value is 0.0274. (Note that by default the `wilcox.test` function in R uses the exact distribution, so the p -value is slightly different. Will discuss exact p -value later.)

- (b) Calculate by hand the Mann-Whitney U-statistic for the comparison of the two groups.

First, there are no ties, so there will be no 1/2’s. Second, in this table, for each observation in group “E”, M_j is the number of “C” subjects that are smaller:

Experimental:	2.8	5.1	7.1	7.7	12.3	18.8	27.1	39.7
M_j	5	6	6	6	7	7	8	8

The sum of the values in the second row is $5 + 6 + 6 + 6 + 7 + 7 + 8 + 8 = 53$. The expected value is $8 \times 8/2 = 32$. Hence $U = 53 - 32 = 21$, the same as the Wilcoxon rank sum. The variance is the same as above so the rest of the computation is identical.

- (c) Using your software of choice (or by hand if you wish), perform the t-test for the comparisons between the two groups.

Using the `t.test` function in R,

```
> t.test(c(0.2,0.8,1.9,2.2,2.6,3.9,8.2,21.8),
+ c(2.8,5.1,7.1,7.7,12.3,18.8,27.1,39.7))
```

Welch Two Sample t-test

data: c(0.2, 0.8, 1.9, 2.2, 2.6, 3.9, 8.2, 21.8) and

```

      c(2.8, 5.1, 7.1, 7.7, 12.3, 18.8, 27.1, 39.7)
t = -1.9093, df = 10.993, p-value = 0.08265
alternative hypothesis: true difference in means is not equal to 0

```

The p -value from the t -test is 0.0827.

Note

- the data do not appear to be close to normal: there are many small values (less than, say, 5 or 10) but several relatively large values (greater than 20). The data are skewed significantly to the right.
- The p -value using the Wilcoxon/Mann-Whitney test is much smaller than the p -value from the t -test.
- For non-normal data, the Wilcoxon/Mann-Whitney test can have greater power than the t -test, which is sensitive to large values. Because the Wilcoxon/Mann-Whitney test is based on ranks, it is insensitive to major deviations from normality.

-
2. Suppose all subjects in a trial are followed for 1 year, and at the end of that time they either survive disease free (DFS), survive but experience a recurrence of disease, or die. Note that these responses are naturally ordered: DFS is good, death is bad, and recurrence is in between. Subjects are randomly assigned either treatment 1 (control) or treatment 2 (experimental). We observe the following:

	DFS	Recurrence	Dead
Treatment 1	33	17	41
Treatment 2	44	18	25

- (a) Compute (by hand) the test statistic for the Wilcoxon rank-sum test assuming that responses are ordered as shown.

The ranks for the DFS subjects range from 1 to 77, for Recurrence from 78 to 112, and the Dead subjects from 113 to 178. Thus the mean ranks in the three groups are $(1+77)/2=39$, $(78+112)/2=95$, and $(113+178)/2=145.5$. The rank sum for Treatment 2 is $T_2 = 44 \times 39 + 18 \times 95 + 25 \times 145.5 = 7063.5$. The expected value of T_2 is $87 \times (178 + 1)/2 = 7785.5$, so $U = 7063.5 - 7785.5 = -723$. The variance is

$$\frac{87 \times 91}{178 - 1} \left(\frac{1}{178} (77 \times 39^2 + 35 \times 95^2 + 66 \times 145.5^2 - \frac{(1 + 178)^2}{4}) \right) = 101620.8$$

The test statistic is $723^2/101620.8 = 5.14$.

-
- (b) What do these results suggest regarding the effect of treatment.

Lower average rank in group 2 suggests that treatment group 2 has better outcomes than treatment group 1.

3. The data file “data2.csv” (in csv format, comma delimited) contains columns

- x_0 : baseline value of response variable
- x_1 : value of response variable at first follow-up time
- x_2 : value of response variable at second follow-up time
- z : treatment variable (0,1)

Assume that the responses are normally distributed.

- (a) For the first follow-up response (x_1) test the null hypothesis that there is no difference by treatment by

- i. ignoring baseline

Fit model for x_1 with just z :

```
> summary(lm(x1~z, data=D))
```

[snip]

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	60.914	1.809	33.677	<2e-16 ***
z	3.886	2.515	1.545	0.128

[snip]

The mean difference is 3.886 with SE 2.515, and t -statistic 1.545.

- ii. using change from baseline ($x_1 - x_0$)

```
> summary(lm(x1-x0~z, data=D))
```

[snip]

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	-0.2679	0.9919	-0.270	0.7881
z	3.5745	1.3792	2.592	0.0122 *

[snip]

The mean difference is 3.5745 with SE 1.3892 and t -statistic 2.592.

- iii. fitting regression model $x_1 = \alpha_0 + \alpha_1 x_0 + \beta z + \epsilon$.

```
> summary(lm(x1~z+x0, data=D))
```

[snip]

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	13.7465	3.6149	3.803	0.000361 ***
z	3.6458	1.2252	2.976	0.004337 **
x_0	0.7709	0.0573	13.454	< 2e-16 ***

[snip]

The mean difference is 3.6458 with SE 1.2252 and t -statistic 2.976.

Why do the conclusions differ from these three analysis?

In (c) the coefficient for x_0 is .771, suggesting (assuming equal variances for x_0 and x_1) that the correlation is greater than .5, so the change from baseline should have smaller variance than the follow-up value alone. This is borne out in the differences between (a) and (b). Since this coefficient is not too close to one, we expect that the regression model in (c) should have smaller variance than either (a) or (b), and again this is borne out in the results.

- (b) Repeat each of (i), (ii), and (iii) above for the response at the second follow-up time (x_2).

```
> summary(lm(x2~z, data=D))
[snip]
Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)   61.618      2.063   29.868  <2e-16 ***
z              6.872      2.868    2.396   0.0200 *
[snip]

> summary(lm(x2-x0~z, data=D))
[snip]
Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)    0.4357      2.6931   0.162   0.8721
z              6.5610      3.7446   1.752   0.0852 .
[snip]

> summary(lm(x2~z+x0, data=D))
[snip]
Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)   52.7430      8.4511   6.241 6.55e-08 ***
z              6.8270      2.8644   2.383   0.0206 *
x0             0.1451      0.1340   1.083   0.2836
[snip]
```

In the third analysis, the coefficient for x_0 is small, suggesting that there is much less correlation between x_2 and x_0 than between x_1 and x_0 . Therefore, we expect that the change from baseline will be much less efficient than ignoring baseline altogether. Again this is borne out in the results. In this case the third analysis gives essentially the same result as the first.

- (c) Comment on the differences between (a) and (b).

Change from baseline beats observed follow-up value alone when correlation between baseline and follow-up is high, and loses when correlation is low. In either case, the regression model is at least as good as the others and should always be preferred.