

Biost 517: Applied Biostatistics I
 #### Emerson, Fall 2005

Annotated Stata Log File: Homework #6
 #### November 28, 2005

In this file I give the Stata commands I used to produce
 #### the key to Homework #6. In order to properly format
 #### a table useful to casual readers, I cut and pasted some
 #### of the output into Excel.

Comments edited into the log file produced by Stata are
 #### on the lines that start with the four '#' signs and are
 #### printed in italics.

The Stata commands are put in **bold face**.

Stata output is displayed in regular typeface in blue.

 #### Reading in the data: I had saved it previously

. use ../../datasets/2005.09.26/dfmowide

NOTE: Recall that Stata has some problem with representing the
 #### dose 0.4 group correctly. Hence, I always selected that group
 #### by asking for dose > 0.3.

 #### Problem 1: Comparing mean spd12 across dose groups using the t test
 #### I chose to use the version of the t test which allows
 #### unequal variances across dose groups.
 #####

Generate descriptive statistics which might be useful in looking for individual
 #### toxicities, etc. with respect to the absolute spermidine levels
. tabstat spd12 if dose==0 | dose>0.3, by(dose) stat(n mean sd min p25 p50 p75
> max) col(stat) format(%9.2f)

Summary for variables: spd12
 by categories of: dose

dose	N	mean	sd	min	p25	p50	p75	max
0	28.00	3.26	1.31	1.01	2.26	2.82	4.27	5.91
.4000000	20.00	1.95	0.80	0.00	1.48	1.93	2.46	3.42
Total	48.00	2.71	1.29	0.00	1.98	2.34	3.32	5.91

. ttest spd12 if dose == 0 | dose > 0.3, by(dose) unequal

Two-sample t test with unequal variances

Group	Obs	Mean	Std. Err.	Std. Dev.	[95% Conf. Interval]
0	28	3.255935	.2482656	1.313698	2.746536 3.765334

4000000		20	1.949568	.178642	.7989114	1.575666	2.323471

combined		48	2.711616	.1866961	1.293469	2.336032	3.0872

diff			1.306367	.3058575		.6903532	1.922381

diff = mean(0) - mean(4000000) t = 4.2712
Ho: diff = 0 Satterthwaite's degrees of freedom = 45.0394

Ha: diff < 0 Ha: diff != 0 Ha: diff > 0
Pr(T < t) = 1.0000 Pr(|T| > |t|) = 0.0001 Pr(T > t) = 0.0000

#####

Problem 2: Comparing geometric mean spd12 across dose groups using the t test on
log transformed data and then back transforming the estimates.
I chose to use the version of the t test which allows
unequal variances across dose groups.
#####

First generate the log transformed data. Note that I have to check for data below
the limit of detection. I choose to use the lowest nonzero measurement made for
spermidine at any time point, so I examine the minimum nonzero value for spd0,
spd6, spd12, and spd15.

. generate lspd12= log(spd12)
(20 missing values generated)

. summ spd0 if spd0 > 0

Variable		Obs	Mean	Std. Dev.	Min	Max

spd0		114	3.409728	1.55291	.66	7.6

. summ spd6 if spd6 > 0

Variable		Obs	Mean	Std. Dev.	Min	Max

spd6		106	2.842533	1.412813	1.05862	7.844

. sum spd12 if spd12 > 0

Variable		Obs	Mean	Std. Dev.	Min	Max

spd12		94	2.798014	1.20635	.293	6.454

. summ spd15 if spd15 > 0

Variable		Obs	Mean	Std. Dev.	Min	Max

spd15		91	2.863014	.8800331	1.249	4.832

. replace lspd12= log(.293/2) if spd12==0
(1 real change made)

Now do the t test which allows unequal variances

. ttest lspd12 if dose==0 | dose > 0.3, by(dose) unequal

Two-sample t test with unequal variances

Group	Obs	Mean	Std. Err.	Std. Dev.	[95% Conf. Interval]	
0	28	1.101382	.077731	.4113139	.9418907	1.260872
4000000	20	.5372735	.1487163	.6650795	.2260067	.8485402
combined	48	.8663365	.0859982	.5958128	.6933305	1.039343
diff		.5641081	.1678054		.2210412	.9071749

```
diff = mean(0) - mean(4000000)                                t =    3.3617  
Ho: diff = 0                Satterthwaite's degrees of freedom = 29.2626
```

Ha: diff < 0	Ha: diff != 0	Ha: diff > 0
Pr(T < t) = 0.9989	Pr(T > t) = 0.0022	Pr(T > t) = 0.0011

```
#### Now exponentiate the estimates for the differences. This will provide the
#### ratio of geometric means. Note that the differences were computed as
#### dose 0 group minus dose 0.4 group. I would have rather had them the other
#### way. So I exponentiated the negative of the estimates. (I could have
#### exponentiated the estimates and then taken the reciprocal, instead.)
```

```
. display exp(-.5641081)
.56886729
```

```
. display exp(-.2210412)
.80168365
```

```
. display exp(-.9071749)
.403663
```

```
#### If I want the geometric means for each dose group separately, I could either
#### exponentiate the values given in the above table, or I could ask Stata to
#### do it for me by using the "means" command by dose.
```

```
. replace spd12= .293/2 if spd12==0
(1 real change made)
```

```
. bysort dose: means spd12
```

-> dose = 0

Variable	Type	Obs	Mean	[95% Conf. Interval]
spd12	Arithmetic	28	3.255935	2.746536 3.765334
	Geometric	28	3.008319	2.564826 3.528499
	Harmonic	28	2.765647	2.351809 3.356228

```
-> dose = .075
```

Variable	Type	Obs	Mean	[95% Conf. Interval]
spd12	Arithmetic	26	2.919683	2.518294 3.321073
	Geometric	26	2.749752	2.375747 3.182634
	Harmonic	26	2.576789	2.23116 3.04913

```
-> dose = .2
```

Variable	Type	Obs	Mean	[95% Conf. Interval]	
spd12	Arithmetic	21	2.71162	2.076408	3.346832
	Geometric	21	2.3247	1.7371	3.111065
	Harmonic	21	1.766532	1.144841	3.865809

```
-> dose = .4
```

Variable	Type	Obs	Mean	[95% Conf. Interval]	
spd12	Arithmetic	20	1.956893	1.591582	2.322205
	Geometric	20	1.711334	1.253584	2.336234
	Harmonic	20	1.167577	.6580205	5.174928

```
#####
```

```
#### Problem 3: Comparing median spd12 across dose groups using bootstrapped SEs.
```

```
#### Note that I have to do this on a "clean" dataset that only contains
```

```
#### cases that I want resampled. So I have to get rid of cases that
```

```
#### have missing data or are in the wrong dose group.
```

```
#####
```

```
#### Restricting the data to the dose 0.4 group with nonmissing data.
```

```
. clear
```

```
. use ../../datasets/2005.09.26/dfmowide
```

```
. drop if spd12==. | dose < 0.4
```

```
(94 observations deleted)
```

```
#### Bootstrapping to find the standard error of the sample median.
```

```
. bs "summ spd12, detail" "_result(10)", reps(1000)
```

```
command:      summ spd12 , detail
statistic:    _bs_1      = _result(10)
```

```
Bootstrap statistics                                Number of obs    =        20
                                                    Replications      =       1000
```

Variable	Reps	Observed	Bias	Std. Err.	[95% Conf. Interval]	
_bs_1	1000	1.929	.0135799	.1666302	1.602015	2.255985 (N)
					1.676	2.279 (P)
					1.7035	2.3 (BC)

```
Note:  N    = normal
       P    = percentile
       BC   = bias-corrected
```

```
. centile spd12
```

```
-- Binom. Interp. --
```

Variable	Obs	Percentile	Centile	[95% Conf. Interval]	
spdl2	20	50	1.929	1.682406	2.295108

```
#### Restricting the data to the dose 0 group with nonmissing data. Note that I had  
#### to drop the data in memory, read in the whole data set, and then drop the  
#### data which is not desired
```

```
. clear
```

```
. use ..\..\datasets\2005.09.26\dfmowide
```

```
. drop if spdl2==. | dose !=0  
(86 observations deleted)
```

```
#### Bootstrapping to find the standard error of the sample median.
```

```
. bs "summ spd12, detail" "_result(10)", reps(1000)
```

```
command:      summ spd12 , detail
statistic:    _bs_1      = _result(10)
```

```
Bootstrap statistics                                Number of obs    =        28
                                                    Replications      =       1000
```

Variable	Reps	Observed	Bias	Std. Err.	[95% Conf. Interval]
_bs_1	1000	2.8155	.0477733	.292795	2.240936 3.390064 (N)
					2.36225 3.7235 (P)
					2.383 3.828 (BC)

```
Note:  N    = normal
       P    = percentile
       BC   = bias-corrected
```

```
#### Computing the difference of sample medians.
```

```
. display (2.8155 - 1.929)
.8865
```

```
#### Computing the lower and upper bounds of the 95% CI for the difference in medians.
```

```
. display (2.8155 - 1.929) - invnorm(.975) * sqrt( .292795^2 + .1666302^2)
.22620871
```

```
. display (2.8155 - 1.929) + invnorm(.975) * sqrt( .292795^2 + .1666302^2)
1.5467913
```

```
#### Computing the Z statistic testing for no difference in medians.
```

```
. display (2.8155 - 1.929) / sqrt( .292795^2 + .1666302^2)
2.6314266
```

```
#### Computing the two-sided P value testing for no difference in medians.
```

```
. display 2 * norm(-2.6314266)
.00850272
```

```
#####
```

```
#### Problem 4: Performing the Wilcoxon rank sum test on the spermidine measurements
#### at 12 months for the dose 0 and dose 0.4 groups.
```

```
#####
```

```
#### Restoring the dataset.
```

```
. clear
```

```
. use ../../datasets\2005.09.26\dfmowide
```

```
#### Performing the rank sum test.
```

```
. ranksum spd12 if dose==0 | dose > 0.3, by(dose)
```

```
Two-sample Wilcoxon rank-sum (Mann-Whitney) test
```

dose	obs	rank sum	expected
0	28	862.5	686
.40000000596	20	313.5	490

```

-----+-----
      combined |          48          1176          1176

unadjusted variance      2286.67
adjustment for ties      -0.12
-----
adjusted variance      2286.54

Ho: spdl2(dose==0) = spdl2(dose==.4000000059604645)
      z =      3.691
      Prob > |z| =      0.0002

```

```
#####
```

```

#### Problem 5: Comparing dose 0 and dose 0.4 groups with respect to the mean change
#### in spermidine measurements over 12 months of treatment using the t test which
#### allows unequal variances on a new variable created to contain the change in
#### spermidine for each subject.
#####

```

```
#### Generating the new variable.
```

```

. g diffspd= spdl2 - spd0
(19 missing values generated)

```

```

#### Generate descriptive statistics which might be useful in looking for individual
#### toxicities, etc. with respect to the change (difference) in spermidine levels

```

```

. tabstat diffspd if dose==0 | dose>0.3, by(dose) stat(n mean sd min p25 p50 p75
> max) col(stat) format(%9.2f)

```

```

Summary for variables: diffspd
      by categories of: dose

```

dose	N	mean	sd	min	p25	p50	p75	max
0	28.00	-0.04	1.53	-3.20	-1.28	0.34	1.12	2.53
.4000000	20.00	-1.76	2.17	-5.48	-3.58	-1.08	-0.24	1.46
Total	48.00	-0.76	2.00	-5.48	-2.05	-0.68	0.80	2.53

```
#### Performing the t test.
```

```

. ttest diffspd if dose==0 | dose > 0.3, by(dose) unequal

```

```
Two-sample t test with unequal variances
```

Group	Obs	Mean	Std. Err.	Std. Dev.	[95% Conf. Interval]	
0	28	-.0408007	.2896714	1.532797	-.6351573	.5535559
4000000	20	-1.755509	.4855573	2.171478	-2.771792	-.7392261
combined	48	-.7552626	.2881597	1.996429	-1.334965	-.17556
diff		1.714709	.5653984		.5631312	2.866286

```

      diff = mean(0) - mean(4000000)          t =      3.0327
Ho: diff = 0          Satterthwaite's degrees of freedom = 32.0722

```

```
Ha: diff < 0
```

```
Ha: diff != 0
```

```
Ha: diff > 0
```

Pr(T < t) = 0.9976

Pr(|T| > |t|) = 0.0048

Pr(T > t) = 0.0024

#####

Problem 6: Comparing dose 0 and dose 0.4 groups with respect to the geometric mean
 #### of the proportionate change in spermidine measurements over 12 months of treatment.
 #### Note that the proportionate change (ratio) is the convenient way to measure
 #### the within individual change when using geometric means. This is easily effected
 #### by first transforming the spd0 and spd12 measurements to their logs, then taking
 #### the within individual difference of those logs (which will be the log of the
 #### within individual ratios), and then using the t test which allows unequal variances.
 #### I then back transform (exponentiate) the estimated differences in log transformed
 #### values to obtain the geometric means.
 #####

Generate descriptive statistics which might be useful in looking for individual
 #### toxicities, etc. with respect to the change (ratio) of spermidine levels

. generate ratspd = spd12 / spd0

. tabstat ratspd if dose==0 | dose>0.3, by(dose) stat(n mean sd min p25 p50 p75
 > max) col(stat) format(%9.2f)

Summary for variables: ratspd
 by categories of: dose

dose	N	mean	sd	min	p25	p50	p75	max
0	28.00	1.10	0.47	0.44	0.70	1.14	1.43	1.89
.4000000	20.00	0.76	0.67	0.00	0.29	0.61	0.89	2.88
Total	48.00	0.96	0.58	0.00	0.53	0.81	1.41	2.88

Generating the new variables.

. generate lspd12= log(spd12)
 (20 missing values generated)

. replace lspd12= log(.293/2) if spd12==0
 (1 real change made)

. generate lspd0= log(spd0)

. replace lspd0= log(.293/2) if spd0==0
 (0 real changes made)

. generate diff12spd= lspd12 - lspd0
 (19 missing values generated)

Computing the t test.

. ttest diff12spd if dose==0 | dose > 0.3, by(dose) unequal

Two-sample t test with unequal variances

Group	Obs	Mean	Std. Err.	Std. Dev.	[95% Conf. Interval]	
0	28	.0016925	.0867118	.4588356	-.1762254	.1796104
4000000	20	-.6304515	.2112218	.9446128	-1.072544	-.188359
combined	48	-.2617008	.1100047	.7621352	-.4830018	-.0403998

```

-----+-----
diff |          .632144    .2283278          .1623018    1.101986
-----+-----
diff = mean(0) - mean(4000000)          t =    2.7686
Ho: diff = 0          Satterthwaite's degrees of freedom = 25.4354

Ha: diff < 0          Ha: diff != 0          Ha: diff > 0
Pr(T < t) = 0.9948    Pr(|T| > |t|) = 0.0104    Pr(T > t) = 0.0052

#### Exponentiating estimates to obtain the ratios of geometric means.
#### For the dose 0 group over 12 months
. display exp(.0016925)
1.0016939

. display exp(-.1762254)
.83842898

. display exp(.1796104)
1.196751

#### For the dose 0.4 group over 12 months
. display exp(-.6304515)
.53235139

. display exp(-1.072544)
.34213701

. display exp(-.0403998)
.96040539

#### The comparison of the two dose groups
. display exp(-.632144)
.53145115

. display exp(-.1623018)
.85018458

. display exp(-1.101986)
.33221066

#####
#### Problem 7: Comparing median change (difference: spd12-spd0) across dose groups
#### using bootstrapped SEs. Note again the need for "clean" data sets.
#####

#### Restricting the data to the dose 0.0 group with nonmissing data.
. g diffspd= spd12 - spd0
(19 missing values generated)

. drop if diffspd==. | dose != 0.0
(86 observations deleted)

. bs "summ diffspd, detail" "_result(10)", reps(1000)

command:    summ diffspd , detail
statistic:  _bs_1      = _result(10)

```

```
Bootstrap statistics                                Number of obs    =        28
                                                    Replications    =       1000
```

Variable	Reps	Observed	Bias	Std. Err.	[95% Conf. Interval]		
_bs_1	1000	.3377751	-.1981491	.508288	-.6596596	1.33521	(N)
					-1.00034	.803605	(P)
					-.87889	.803605	(BC)

```
Note:  N    = normal
       P    = percentile
       BC   = bias-corrected
```

```
#### Restricting the data to the dose 0.4 group with nonmissing data.
```

```
. clear
```

```
. use ..\..\datasets\2005.09.26\dfmowide
```

```
. g diffspd= spd12 - spd0
(19 missing values generated)
```

```
. drop if diffspd==. | dose < 0.3
(94 observations deleted)
```

```
. bs "summ diffspd, detail" "_result(10)", reps(1000)
```

```
command:      summ diffspd , detail
statistic:    _bs_1      = _result(10)
```

```
Bootstrap statistics                                Number of obs    =        20
                                                    Replications    =       1000
```

Variable	Reps	Observed	Bias	Std. Err.	[95% Conf. Interval]		
_bs_1	1000	-1.07527	-.3551135	.7539308	-2.55474	.4041996	(N)
					-3.04967	-.444005	(P)
					-2.81234	-.2420051	(BC)

```
Note:  N    = normal
       P    = percentile
       BC   = bias-corrected
```

```
#### Computing the difference in the sample median of the changes for each group.
```

```
. display -1.07527 - .3377751
-1.4130451
```

```
#### Computing the Z score for a test of no difference in the median change.
```

```
. display (-1.07527 - .3377751) / sqrt(.7539308^2 +.508288^2)
-1.5540468
```

```
#### Computing the two-sided P value for a test of no difference in the median change.
```

```
. display 2* norm(-1.5540468)
.12017325
```

```
#### Computing the 95% CI for the difference in the median change.
. display (-1.07527 - .3377751) - invnorm(.975) * sqrt(.7539308^2 +.508288^2)
-3.1951777
```

```
. display (-1.07527 - .3377751) + invnorm(.975) * sqrt(.7539308^2 +.508288^2)
.36908748
```

```
#####
#### Problem 8: Performing the Wilcoxon rank sum test comparing the dose 0.4 and
#### dose 0 groups with respect to the difference in spermidine
#### measurements over the 12 months of treatment (spd12 - spd0).
#####
```

```
#### Restoring the dataset.
```

```
. clear
```

```
. use ../../datasets\2005.09.26\dfmowide
```

```
. g diffspd= spd12 - spd0
(19 missing values generated)
```

```
. ranksum diffspd if dose==0 | dose > 0.3, by(dose)
```

Two-sample Wilcoxon rank-sum (Mann-Whitney) test

dose	obs	rank sum	expected
0	28	810	686
.40000000596	20	366	490
combined	48	1176	1176

unadjusted variance 2286.67

adjustment for ties 0.00

adjusted variance 2286.67

Ho: diffspd(dose==0) = diffspd(dose==.4000000059604645)

z = 2.593

Prob > |z| = 0.0095

```
#####
#### Problem 9: Performing a chi squared test comparing the dose 0.4 and
#### dose 0 groups with respect to the proportion of individuals
#### who exhibited a decrease in spermidine after 12 months of
#### treatment.
#####
```

```
#### Creating a variable indicating individuals whose spermidine values
#### were observed to decrease. Note the need to consider missing data.
```

```
. g decrspd = 0
```

```
. replace decrspd=1 if diffspd < 0
(58 real changes made)
```

```
. replace decrspd=. if diffspd == .
(19 real changes made, 19 to missing)
```

Inference for proportion with decrease in spd levels by dose group.

. bysort dose: ci decrspd

-> dose = 0

Variable	Obs	Mean	Std. Err.	[95% Conf. Interval]	
-----+-----					
decrspd	28	.4642857	.0959793	.2673525	.6612189

-> dose = .075

Variable	Obs	Mean	Std. Err.	[95% Conf. Interval]	
-----+-----					
decrspd	26	.6153846	.0973009	.4149898	.8157795

-> dose = .2

Variable	Obs	Mean	Std. Err.	[95% Conf. Interval]	
-----+-----					
decrspd	21	.6190476	.1085881	.3925367	.8455585

-> dose = .4

Variable	Obs	Mean	Std. Err.	[95% Conf. Interval]	
-----+-----					
decrspd	20	.8	.0917663	.6079309	.9920691

Performing the chi squared test. Note how the Stata labeling of the
 #### output is nearly incomprehensible. "Unexposed" is the dose 0 group,
 #### and "Exposed" is the group with dose not equal to zero. "Cases"
 #### will be those individuals who had a 1 for the decrspd variable.

. cs decrspd dose if dose==0 | dose > 0.3

	dose		
	Exposed	Unexposed	Total
-----+-----			
Cases	16	13	29
Noncases	4	15	19
-----+-----			
Total	20	28	48
Risk	.8	.4642857	.6041667
	Point estimate		[95% Conf. Interval]
-----+-----			
Risk difference	.3357143		.081047 .5903816
Risk ratio	1.723077		1.094051 2.713763
Attr. frac. ex.	.4196429		.0859654 .631508
Attr. frac. pop	.2315271		
-----+-----			
	chi2(1) =		5.50 Pr>chi2 = 0.0190