

Homework 6, due in due Thursday Oct. 20 in lab **Solutions**

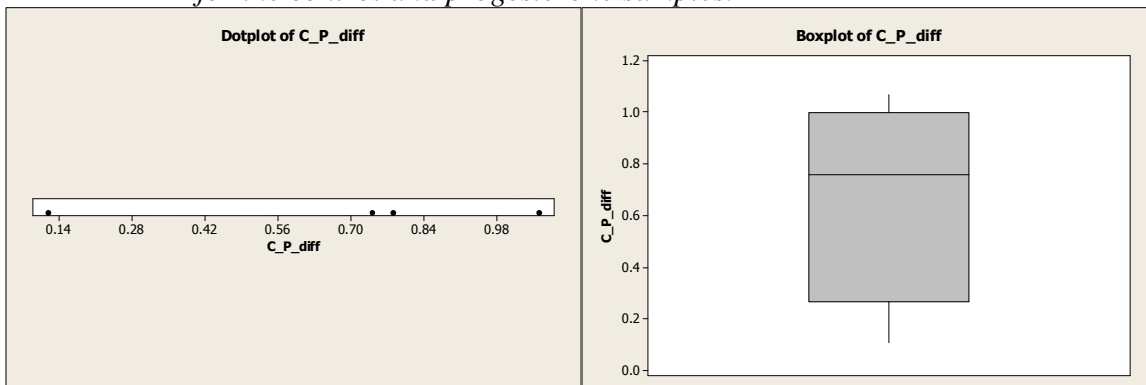
(assignment for these solutions can be found on the last page)

1. *Cyclic adenosine monophosphate (cAMP) is a substance that can mediate cellular response to hormones. In a study of maturation of egg cells in the frog *Xenopus laevis*, oocytes from each of four females were divided into two batches: one batch was exposed to progesterone and the other was not. After two minutes, each batch was assayed for its cAMP content, with the results given in the table below.*

- a. *Explain why this is a paired design.*

The two populations (treatment and control) are paired in such a way that they were as close as possible as the same. This was done by using two oocytes from the same female and assigning one to each population.

- b. *Make a dotplot and box plot of the differences between the cAMP levels for the control and progesterone samples.*



All the differences are positive.

- c. *Test at the 5% level whether there is any difference in the population mean cAMP levels for batches of oocytes that are untreated versus those treated with progesterone.*

The t-test below has a *p*-value of 0.044, which is sufficient evidence to conclude there is a difference in the two population mean cAMP levels. Similarly, the confidence interval (0.031587, 1.318413) does not include 0, the value for no difference.

**Paired T-Test and CI: Control, Progesterone**

Paired T for Control - Progesterone				
	N	Mean	StDev	SE Mean
Control	4	2.98000	2.04706	1.02353
Progesterone	4	2.30500	1.95186	0.97593
Difference	4	0.675000	0.404351	0.202176

95% CI for mean difference: **(0.031587, 1.318413)**

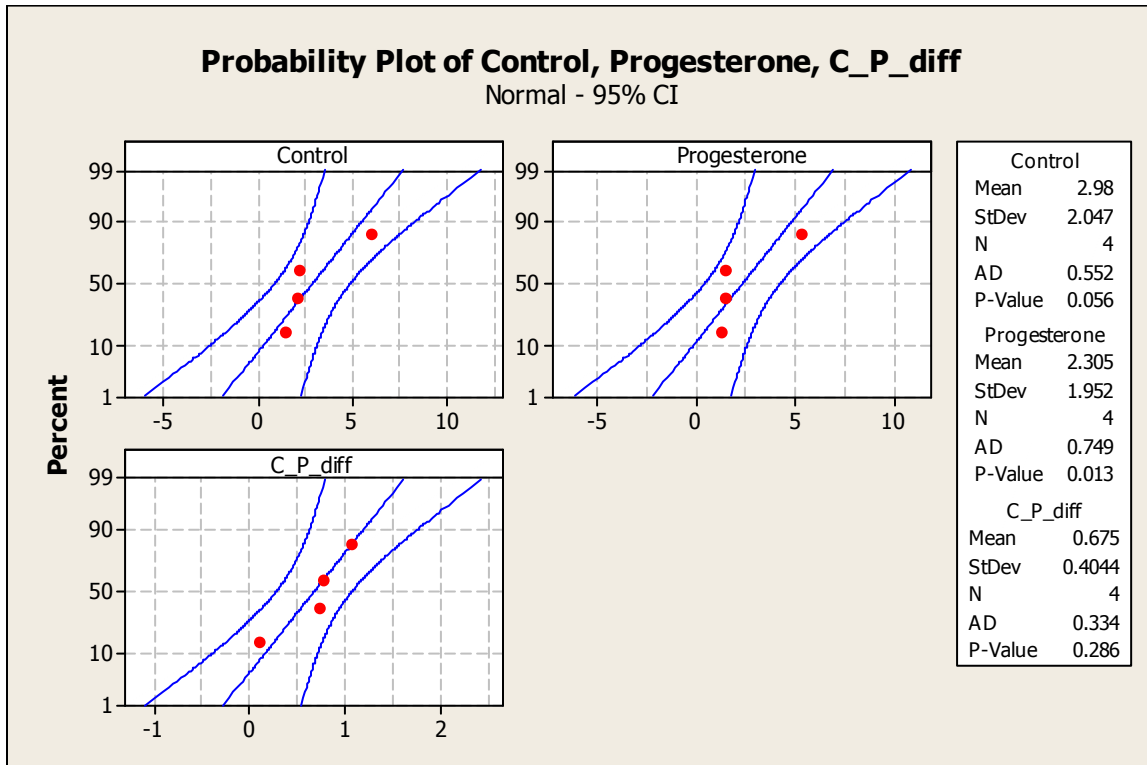
T-Test of mean difference = 0 (vs not = 0): T-Value = 3.34 **P-Value = 0.044**

- d. Obtain and interpret a 95% CI for the difference in population mean cAMP levels for batches of oocytes that are untreated versus those treated with progesterone.

From above, the confidence interval (0.031587, 1.318413) contains the true population difference in means with 95% confidence. That is, 95% of intervals constructed in this way contain the true difference in population means. Because 0 is not included in the interval, we conclude there is a difference in population means.

- e. Discuss any statistical assumptions that you have made in carrying out the analysis, and whether the assumptions seem reasonable.

Because these samples are extremely small (only 4 each), we are assuming the populations are each normal. The probability plots below indicate one observation in each sample that deviates from an otherwise straight line through the other three. The  $p$ -values for the Anderson-Darling normality test are 0.056 and 0.013, which are both evidence against the normality assumption. However, the sample of differences appears reasonably normal with a  $p$ -value=0.286.



- f. Write a short summary to the problem.

To study the difference in cAMP content in frogs when exposed to progesterone, a matched-pairs design was used in a sample of size 4 in each group. A hypothesis test of whether there is a difference was performed, and while all differences were positive, there was evidence to conclude a difference. However, probability plots indicate the populations of the individual samples are not normal, yet our conclusions based on tiny samples of size 4 should be made with caution.

2. For these data, we are interested in comparing the distribution of birth weights for males to females.

a. Using Minitab, describe the shapes, and make sure to compare the centers and spreads for the birth weights of males and females. Do the distributions appear to be reasonably normal with similar spreads? Does there appear to be much of a difference in the typical birth weights for males compared to females? Discuss briefly.

The male data are roughly normal, the female data are skewed right.

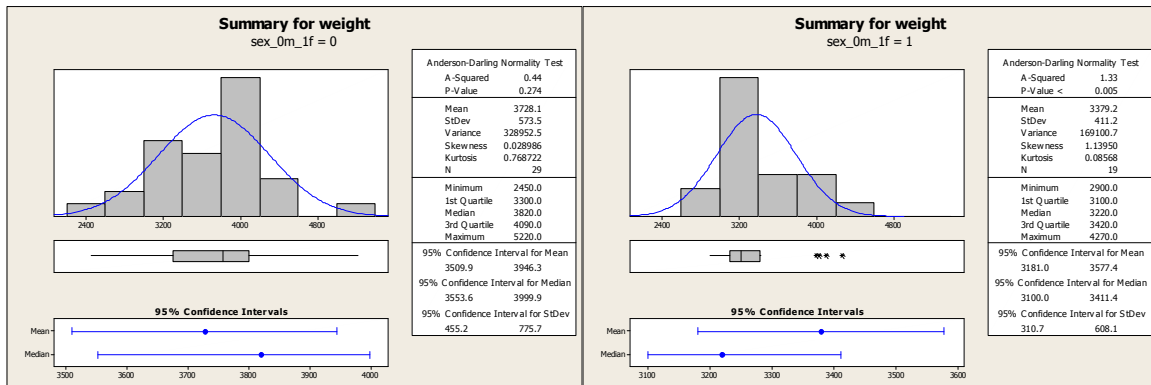
The male data are centered 400-600 grams greater based on mean and median than the female data.

The male spread is about 25% greater in terms of standard deviation, and twice as large in terms of Range and IQR.

There appears to be a difference since, in terms of SE, the means are roughly 4 units from each other.

**Descriptive Statistics: weight**

sex	N	Mean	SE Mean	StDev	Minimum	Q1	Median	Q3	Maximum	Range	IQR
0	29	3728	107	574	2450	3300	3820	4090	5220	2770	790
1	19	3379.2	94.3	411.2	2900.0	3100.0	3220.0	3420.0	4270.0	1370.0	320.0



b. Do a formal test at the 5% level that the population mean birth weights are the same for males and females. Use the pooled t-test and do the test by comparing the p-value to .05. Also, compare the result with what you get from the Satterthwaite (Welch) test.

A pooled t-test gives a p-value=0.027, which at the 5% level causes us to reject H0 in favor of HA, concluding the population means are different from each other. Using nonequal variances, we have slightly stronger evidence for a difference in population means, with a p-value=0.018.

**Two-Sample T-Test and CI: weight, sex\_0m\_1f**

Two-sample T for weight

sex_0m_1f	N	Mean	StDev	SE Mean
0	29	3728	574	107
1	19	3379	411	94

Difference = mu (0) - mu (1)  
Estimate for difference: 348.893  
95% CI for difference: (42.249, 655.537)

T-Test of difference = 0 (vs not =): T-Value = 2.29 **P-Value = 0.027** DF = 46  
Both use Pooled StDev = 516.1413

### Unpooled

95% CI for difference: **(62.329, 635.457)**

T-Test of difference = 0 (vs not =): T-Value = 2.45 **P-Value = 0.018** DF = 45

- c. Construct and interpret a 95% CI for the difference in population mean birth weights between males and females.*

The confidence intervals above do not contain 0. 95% of intervals constructed in this way contain the true difference in population means.

- d. Write a short summary of your analysis. Include any problems with assumptions you might see that could affect the validity of your conclusions.*

In a test of the difference in population means between boy and girl weights, we find at a 5% significance level that there is a difference.

The sample of girl weights is not normal, and with a small sample this violates the assumptions of the hypothesis test and confidence intervals.

- e. Extra Credit: Repeat parts b and c using the Mann-Whitney test. Would you change any conclusions using that analysis?*

The test for a difference in population medians concludes a difference at a 5% level ( $p$ -value=0.0341). The confidence interval is wider than above, and does not include 0. The conclusion is the same.

### Mann-Whitney Test and CI: weight\_0, weight\_1

	N	Median
weight_0	29	3820.0
weight_1	19	3220.0

Point estimate for ETA1-ETA2 is 430.0

95.0 Percent CI for ETA1-ETA2 is **(20.1, 699.9)**

W = 811.5

Test of ETA1 = ETA2 vs ETA1 not = ETA2 is significant at **0.0341**

The test is significant at **0.0340** (adjusted for ties)

Stat 538: Biostatistics I Homework 6, due Thursday Oct. 20 in lab

1. Cyclic adenosine monophosphate (cAMP) is a substance that can mediate cellular response to hormones. In a study of maturation of egg cells in the frog *Xenopus laevis*, oocytes from each of four females were divided into two batches: one batch was exposed to progesterone and the other was not. After two minutes, each batch was assayed for its cAMP content, with the results given in the table below.

Frog	cAMP (pmol/oocyte)	
	Control	Progesterone
1	6.01	5.23
2	2.28	1.21
3	1.51	1.40
4	2.12	1.38

Use Minitab for this problem.

- Explain why this is a paired design.
- Make a dotplot and box plot of the differences between the cAMP levels for the control and progesterone samples.
- Test at the 5% level whether there is any difference in the population mean cAMP levels for batches of oocytes that are untreated versus those treated with progesterone.
- Obtain and interpret a 95% CI for the difference in population mean cAMP levels for batches of oocytes that are untreated versus those treated with progesterone.
- Discuss any statistical assumptions that you have made in carrying out the analysis, and whether the assumptions seem reasonable.
- Write a short summary to the problem.

2. These data were used in HW 3 (and appear again below). Do all the calculations in Minitab.

The following data were collected from 48 women who were at least 40 years old when they gave birth to their first child. The data concern the gestation period of that pregnancy, and related variables on the child and mother.

The columns are, from left to right:

- ID
- The child's gestation period, in weeks
- Sex of the child (0=Male, 1=Female)
- Birth Weight of child, in grams
- Number of cigarettes smoked per day (on average) by the mother
- Height of mother in cm
- Weight of mother in kilograms at first prenatal visit
- Weight of mother in kilograms at final prenatal visit

For these data, we are interested in comparing the distribution of birth weights for males to females.

- Using Minitab, describe the shapes, and make sure to compare the centers and spreads for the birth weights of males and females. Do the distributions appear to be reasonably normal with similar spreads? Does there appear to be much of a difference in the typical birth weights for males compared to females? Discuss briefly.
- Do a formal test at the 5% level that the population mean birth weights are the same for males and females. Use the pooled t-test and do the test by comparing the p-value to .05. Also, compare the result with what you get from the Satterthwaite (Welch) test.
- Construct and interpret a 95% CI for the difference in population

Stat 538 – Biostatistics I  
Homework 6 solutions

mean birth weights between males and females.

d. Write a short summary of your analysis. Include any problems with assumptions you might see that could affect the validity of your conclusions.

Extra Credit: Repeat parts b and c using the Mann-Whitney test. Would you change any conclusions using that analysis?

DATA

----

01	36	0	3300	0	160.0	67.3	82.7
02	38	0	3300	60	167.6	52.7	76.0
03	38	0	4100	20	167.6	64.2	79.6
04	38	1	2900	10	163.9	72.7	95.8
05	39	0	2820	0	161.3	50.0	63.3
06	39	0	3040	0	158.8	49.1	61.5
07	39	0	4120	0	160.0	57.7	73.5
08	39	0	4200	0	174.0	68.0	86.8
09	39	1	3100	0	171.5	67.3	85.6
10	39	1	3330	0	160.0	74.0	90.5
11	39	1	3410	0	165.1	55.9	70.7
12	39	1	3420	0	162.6	52.3	66.0
13	40	0	2450	20	167.6	61.4	72.5
14	40	0	2885	0	167.7	60.0	78.6
15	40	0	3235	0	170.2	50.0	65.5
16	40	0	3320	0	165.1	63.6	80.2
17	40	0	3600	0	165.1	53.2	68.7
18	40	0	3720	0	165.0	57.7	74.4
19	40	0	3720	0	172.7	61.4	80.0
20	40	0	3820	0	175.3	60.8	78.1
21	40	0	3840	0	167.0	60.5	83.9
22	40	0	3880	0	156.2	57.3	73.7
23	40	0	3960	0	157.5	52.7	68.2
24	40	0	4465	0	157.5	51.4	66.4
25	40	1	2980	0	160.0	47.7	55.2
26	40	1	3040	0	162.0	49.0	60.3
27	40	1	3060	20	157.5	61.0	75.0
28	40	1	3100	0	170.2	55.5	64.6
29	40	1	3120	0	160.3	56.8	75.4
30	40	1	3205	0	172.7	58.2	75.5
31	40	1	3220	0	170.0	64.6	86.0
32	40	1	4100	40	167.0	67.0	85.0
33	41	0	3100	0	168.9	61.4	69.2
34	41	0	3720	0	170.2	57.7	67.7
35	41	0	3720	20	170.2	57.7	80.5
36	41	0	3900	0	167.0	68.0	85.4
37	41	0	3990	0	165.1	52.3	71.2
38	41	0	4050	0	167.6	61.0	78.5
39	41	0	4080	0	162.6	59.1	83.1
40	41	0	4100	0	165.1	60.5	86.5
41	41	0	4460	20	165.1	56.8	88.0
42	41	0	5220	0	157.5	56.8	68.2
43	41	1	3300	40	162.6	74.1	89.7
44	41	1	3400	0	172.7	71.4	87.8
45	41	1	4000	0	165.1	90.0	100.8
46	41	1	4030	0	166.0	63.0	95.3
47	43	1	3220	0	166.4	60.9	72.0
48	43	1	4270	0	162.6	54.5	70.3