Practice Exam Questions and Solutions for the Final Exam; Fall, 2008
Statistics 301, Professor Wardrop
Part A, Chapters 7, 12 and 15

Chapter 7

1. An observational study yields the following "collapsed table."

Group	S	F	Total
1	72	228	300
2	88	212	300
Total	160	440	600

Below are two component tables for these data. Complete these tables so that Simpson's Paradox is occurring **or** explain why Simpson's Paradox *cannot* occur for these data. For the latter, you must provide computations that justify your answer.

Subgp A				Subgp B			
Gp	S	F	Tot	Gp	S	F	Tot
1	30	30	60	1	42	198	240
2			120	2			180
Tot			180	Tot			420

2. An observational study yields the following "collapsed table."

Group	S	F	Total
1	100	100	200
2	156	244	400
Total	256	344	600

Below are two component tables for these data. Complete these tables so that Simpson's Paradox is occurring **or** explain why Simpson's Paradox *cannot* occur for these data. For the latter, you must provide computations that justify your answer.

Subgp A				Subgp B			
Gp	S	F	Tot	Gp	S	F	Tot
1	70	30	100	1	30	70	100
2			80	2			320
Tot			180	Tot			420

3. An observational study yields the following collapsed table.

Group	S	F	Total
1	41	59	100
2	35	65	100
Total	76	124	200

Below are two component tables for these data. Complete these tables so that Simpson's Paradox is occurring **or** explain why Simpson's Paradox *cannot* occur for these data. For the latter, you must provide computations that justify your answer.

Subgp A				Subgp B			
Gp	S	F	Tot	Gp	S	F	Tot
1	5	15	20	1	36	44	80
2			40	2			60
Tot			60	Tot			140

4. An observational study yields the following collapsed table.

Group	S	F	Total
1	100	150	250
2	42	58	100
Total	142	208	350

Below are two component tables for these data. Complete these tables so that Simpson's Paradox is occurring **or** explain why Simpson's Paradox *cannot* occur for these data. For the latter, you must provide computations that justify your answer.

Subgp A				Subgp B			
Gp	S	F	Tot	Gp	S	F	Tot
1	57	93	150	1	43	57	100
2			60	2			40
Tot			210	Tot			140

5. In a 'Chapter 7' problem, you are given the following information:

$$\sqrt{\frac{\hat{p}_1\hat{q}_1}{n_1}} = 0.025, \text{ and}$$

$$\frac{\hat{p}_2\hat{q}_2}{n_2} = 0.005.$$

Calculate the half-width of the 90% confidence interval for $p_1 - p_2$.

6. In a 'Chapter 7' problem, you are given the following information:

$$1.96\sqrt{\frac{\hat{p}_1\hat{q}_1}{n_1}} = 0.0679, \text{ and}$$

$$1.645\sqrt{\frac{\hat{p}_2\hat{q}_2}{n_2}} = 0.0409.$$

Calculate the half-width of the 80% confidence interval for $p_1 - p_2$.

Chapter 12

7. A sample of size 40 yields the following sorted data. Note that I have deleted $x_{(39)}$ (the second largest number). This fact will NOT prevent you from answering the questions below.

14.1	46.0	49.3	53.0	54.2	54.7	54.7
54.7	54.8	55.4	57.6	58.2	58.3	58.7
58.9	60.8	60.9	61.0	61.1	63.0	64.3
65.6	66.3	66.6	67.0	67.9	70.1	70.3
72.1	72.4	72.9	73.5	74.2	75.3	75.4
75.9	76.5	77.0	$x_{(39)}$	88.9		

(a) Construct the density scale histogram for these data using the class intervals: $[10, 40]$, $[40, 50]$, $[50, 60]$, $[60, 75]$ and $[75, 90]$.

(b) Calculate the first quartile and median of these data.

(c) Given that the mean of these data is 63.50 and the standard deviation is 12.33, what proportion of the data lie within one standard deviation of the mean?

(d) Suppose that we learn that 14.1 is a misprint. The actual observation is 64.1. Recalculate the mean after changing the 14.1 to 64.1. Also, recalculate the first quartile and median.

8. A sample of size 50 yields the following sorted data.

0.01	0.04	0.04	0.06	0.08	0.10	0.12
0.14	0.23	0.27	0.29	0.39	0.44	0.48
0.50	0.58	0.58	0.62	0.70	0.71	0.72
0.73	0.76	0.78	0.88	1.00	1.08	1.09
1.10	1.10	1.14	1.16	1.30	1.30	1.31
1.57	1.60	1.74	1.80	1.85	1.88	1.96
2.22	2.27	2.45	2.96	3.44	3.86	4.05
4.52						

Hint: For these data, $\bar{x} = 1.20$ and $s = 1.09$.

(a) Construct the density scale histogram for these data using the class intervals: $[0.00, 0.50]$, $[0.50, 1.00]$, $[1.00, 2.00]$ and $[2.00, 5.00]$.

(b) Calculate the median of these data.

(c) What proportion of these data lie within one standard deviation of the mean?

(d) Calculate Q_1, Q_3 and the mean after deleting the two largest observations.

Chapter 15

9. Independent random samples are selected from two populations. Below are the sorted data from the first population.

362	373	399	428	476	481
545	564	585	589	590	600
671	694	723	724	904	

Hint: The mean and standard deviation of these numbers are 571.1 and 144.7.

Below are the sorted data from the second population.

387	530	544	547	646	766
786	864				

Hint: The mean and standard deviation of these numbers are 633.8 and 160.8.

(a) Calculate Gosset's 90% confidence interval for the mean of the first population.

(b) Calculate a confidence interval for the median of the second population. Select your confidence level and report it with your answer.

(c) Suppose that we now learn that the two samples came from the same population. Thus, the two samples can be combined into one random sample from the one population. Use this combined sample to obtain the 95% confidence interval for the median of the population.

10. Independent random samples are selected from two populations. Below are the sorted data from the first population.

61.6	64.7	65.2	67.2	68.4	69.8
72.5	73.4	76.1	76.5	77.2	78.5
80.0	81.1	84.1	86.6	90.4	91.0

Hint: The mean and standard deviation of these numbers are 75.79 and 8.76.

Below are the data from the second population.

69.4	67.6	74.2	72.1	74.0	67.9
78.7	83.6				

Hint: The mean and standard deviation of these numbers are 73.44 and 5.54.

(a) Calculate Gosset's 98% confidence interval for the mean of the first population.

(b) Calculate a confidence interval for the median of the second population. Select your confidence level and report it with your answer.

(c) Suppose that we now learn that the two samples came from the same population. Thus, the two samples can be combined into one random sample from the one population. Use this combined sample to obtain the 80% confidence interval for the median of the population.

11. Recall that a confidence interval is *too small* if the number being estimated is *larger* than every number in the confidence interval. Similarly, a confidence interval is *too large* if the number being estimated is *smaller* than every number in the confidence interval.

Each of four researchers selects a random sample from the same population. Each researcher calculates a confidence interval for the mean of the population. The intervals are below.

$$[10, 52], [24, 69], [37, 62] \text{ and } [58, 81].$$

(a) Nature announces, "Two of the intervals are correct, one interval is too small and one interval is too large." Given this information, determine all possible values of the mean. **Do not explain.**

(b) Nature announces, "Exactly three of the intervals are correct." Given this information, determine all possible values of the mean. **Do not explain.**

12. Recall that a confidence interval is *too small* if the number being estimated is *larger* than every number in the confidence interval. Similarly, a confidence interval is *too large* if the number being estimated is *smaller* than every number in the confidence interval.

Each of four researchers selects a random sample from the same population. Each researcher calculates a confidence interval for the mean of the population. The intervals are below.

$$[43, 82], [49, 92], [57, 95] \text{ and } [86, 112].$$

(a) Nature announces, "Two of the intervals are correct, one interval is too small and one interval is too large." Given this information, determine all possible values of the mean. **Do not explain.**

(b) Nature announces, "Exactly three of the intervals are correct." Given this information, determine all possible values of the mean. **Do not explain.**

13. Eighteen researchers select random samples from the same population. Each researcher computes a confidence interval for the mean of the population. Thus, each researcher is estimating the same number. Unfortunately, the lower and upper bounds of the 18 confidence intervals became disconnected. The 18 lower bounds, sorted, are below.

153	163	188	198	245	246
260	284	288	307	308	322
334	344	360	375	393	409

The 18 upper bounds, sorted, are below.

294	326	331	345	351	364
372	373	381	394	396	401
403	408	411	439	445	460

If you don't remember what it means for a confidence interval to be 'too small' or 'too large,' see question 11.

(a) Given that $\mu = 300$, how many of the confidence intervals are correct?

(b) Suppose we are told that exactly two of the intervals are too large. Determine all possible values of μ.

(c) Suppose we are told that at least five of the intervals are too small. Determine all possible values of μ.

(d) It could happen that all of the confidence intervals are incorrect; for example, if $\mu = 0$ then one can conclude that all of the intervals are too large. It can also be shown that it is impossible for all 18 intervals to be correct. Here is the two part question for you:

> What is the largest number of these intervals that could possibly be correct? Which values of μ will achieve this maximum number of correct intervals?

14. Twenty-four researchers select random samples from the same population. Each researcher computes a confidence interval for the mean of the population. Thus, each researcher is estimating the same number. Unfortunately, the lower and upper bounds of the 24 confidence intervals became disconnected. The 24 lower bounds, sorted, are below.

328	465	501	513	550	560
561	564	577	580	581	605
610	633	634	639	642	646
649	695	720	756	794	802

The 24 upper bounds, sorted, are below.

518	535	565	568	592	598
612	645	655	656	661	666
688	693	701	752	759	772
781	806	817	837	862	873

If you don't remember what it means for a confidence interval to be 'too small' or 'too large,' see question 12.

(a) Given that $\mu = 600$, how many of the confidence intervals are correct?

(b) Suppose we are told that exactly three of the intervals are too small. Determine all possible values of μ.

(c) Suppose we are told that at least four of the intervals are too large. Determine all possible values of μ.

(d) It could happen that all of the confidence intervals are incorrect; for example, if $\mu = 0$ then one can conclude that all of the intervals are too large. It can also be shown that it is impossible for all 24 intervals to be correct. Here is the two part question for you:

> What is the largest number of these intervals that could possibly be correct? Which values of μ will achieve this maximum number of correct intervals?

Solutions

1. In the collapsed table, $\hat{p}_1 < \hat{p}_2$. To get a reversal, we need

$$c/120 < 30/60, \text{ or } c < 60 \text{ in Sbgp A and}$$

$$c/180 < 42/240, \text{ or } c < 31.5$$

in Subgroup B. Also, the two c's must sum to the 88 in the collapsed table. There are three possible answers: 59 and 29; 58 and 30; 57 and 31.

2. In the collapsed table, $\hat{p}_1 > \hat{p}_2$. To get a reversal, we need

$$c/80 > 70/100, \text{ or } c > 56 \text{ in Sbgp A and}$$

$$c/320 > 30/100, \text{ or } c > 96$$

in Subgroup B. Also, the two c's must sum to the 156 in the collapsed table. Possible answers are: 57 and 99; 58 and 98; 59 and 97.

3. In the collapsed table, $\hat{p}_1 > \hat{p}_2$. To get a reversal, we need

$$c/40 > 5/20, \text{ or } c > 10 \text{ in Sbgp A and}$$

$$c/60 > 36/80, \text{ or } c > 27$$

in Subgroup B. Also, the two c's must sum to the 35 in the collapsed table. This combination of restrictions is impossible.

4. In the collapsed table, $\hat{p}_1 < \hat{p}_2$. To get a reversal, we need

$$c/60 < 57/150, \text{ or } c < 22.8 \text{ in Sbgp A and}$$

$$c/40 < 43/100, \text{ or } c < 17.2$$

in Subgroup B. Also, the two c's must sum to the 42 in the collapsed table. This combination of restrictions is impossible.

5. The half-width is

$$1.645\sqrt{(0.025)^2 + 0.005} = 0.123.$$

6. First,

$$\frac{\hat{p}_1\hat{q}_1}{n_1} = (\frac{0.0679}{1.96})^2 = 0.0012, \text{ and}$$

$$\frac{\hat{p}_2\hat{q}_2}{n_2} = (\frac{0.0409}{1.645})^2 = 0.0006.$$

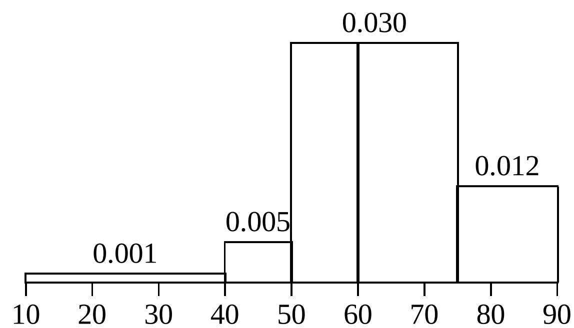
Thus, the half-width is

$$1.282\sqrt{0.0012 + 0.0006} = 0.0544.$$

7. (a) First, we create the following table.

Class Interval	Freq.	Rel. Fr.	Density (rounded)
10–40	1	0.025	0.001
40–50	2	0.050	0.005
50–60	12	0.300	0.030
60–75	18	0.450	0.030
75–90	7	0.175	0.012

The density scale histogram is below.



(b) The median is $(63.0 + 64.3)/2 = 63.65$. The first quartile is $(55.4 + 57.6)/2 = 56.5$.

(c) The one sd interval ranges from

$$63.50 - 12.33 = 51.17, \text{ to}$$

$$63.50 + 12.33 = 75.83.$$

This interval contains 32 observations; thus, the proportion within it is $32/40 = 0.80$.

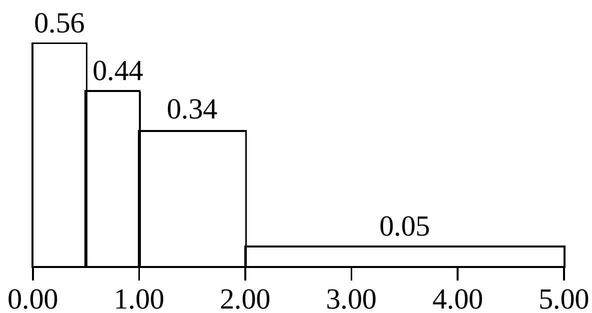
(d) Changing 14.1 to 64.1 increases the total of the 40 numbers by 50. Thus, the mean grows by $50/n = 50/40 = 1.25$. The new mean is 64.75.

The new Q_1 is $(57.6 + 58.2)/2 = 57.9$. The new median is $(64.1 + 64.3)/2 = 64.2$.

8. (a) First, we create the following table.

Class Interval	Freq.	Rel. Fr.	Density (rounded)
0.00–0.50	14	0.28	0.56
0.50–1.00	11	0.22	0.44
1.00–2.00	17	0.34	0.34
2.00–5.00	8	0.16	0.05

The density scale histogram is below.



(b) The median is $(0.88 + 1.00)/2 = 0.94$.

(c) The one sd interval ranges from $1.20 - 1.09 = 0.11$ to $1.20 + 1.09 = 2.29$. This interval contains 38 observations; thus, the proportion within it is $38/50 = 0.76$.

(d) The deletions leave $n = 48$ numbers, a multiple of 4. Thus,

$$Q_1 = (0.39 + 0.44)/2 = 0.415 \text{ and}$$

$$Q_3 = (1.57 + 1.60)/2 = 1.585.$$

For the mean, do not sum the 48 numbers! Instead, the total of the 50 numbers is $50(1.20) = 60$; after deletions, the total of the 48 numbers is $60 - 4.05 - 4.52 = 51.43$. Thus, the mean of the 48 numbers is $51.43/48 = 1.07$.

9. (a) First, $n = 17$; thus, there are 16 degrees of freedom making $t = 1.746$. The CI is

$$571.1 \pm 1.746(144.7)/\sqrt{17} =$$

$$571.1 \pm 61.3 = [509.8, 632.4].$$

(b) B/c $n = 8$ is smaller than 21, we can use the exact CI. From Table A.7 the possible answers are:

- 99.2%: $[387, 864]$, and
- 93.0%: $[530, 786]$.

(c) B/c $n = 25$ is larger than 20, we must use the approximate CI. First,

$$k' = \frac{25+1}{2} - \frac{1.96\sqrt{25}}{2} =$$

$$13 - 4.9 = 8.1.$$

Thus, $k = 8$ and the CI is $[530, 671]$.

10. (a) First, $n = 18$; thus, there are 17 degrees of freedom making $t = 2.567$. The CI is

$$75.79 \pm 2.567(8.76)/\sqrt{18} =$$

$$75.79 \pm 5.30 = [70.49, 81.09].$$

(b) Note that the data are not sorted. B/c $n = 8$ is smaller than 21, we can use the exact CI. From Table A.7 the possible answers are:

- 99.2%: $[67.6, 83.6]$, and
- 93.0%: $[67.9, 78.7]$.

(c) B/c $n = 26$ is larger than 20, we must use the approximate CI. First,

$$k' = \frac{26+1}{2} - \frac{1.282\sqrt{26}}{2} =$$

$$13.5 - 3.27 = 10.23.$$

Thus, $k = 10$ and the CI is $[72.1, 77.2]$.

11. (a) The key is to save the correct intervals for the last. One interval too large means that one lower bound is too large; this must be the 58. Thus, $\mu < 58$. The one interval too small means that one upper bound is too small; this must be the 52. Thus, $\mu > 52$. The remaining intervals must be correct. Thus, $37 \leq \mu \leq 62$.

All three of these conditions must be met, so

$$52 < \mu < 58.$$

(b) Note that the first and last intervals in the list do not overlap, so they cannot both be correct. Thus, the three correct intervals Nature refers to are either: the first three or the last three. If the first three:

$$37 \leq \mu \leq 52;$$

if the last three:

$$58 \leq \mu \leq 62.$$

So the answer is that μ is in the interval $[37, 52]$ or the interval $[58, 62]$.

12. (a) The key is to save the correct intervals for the last. One interval too large means that one lower bound is too large; this must be the 86. Thus, $\mu < 86$. The one interval too small means that one upper bound is too small; this must be the 82. Thus, $\mu > 82$. The remaining intervals must be correct. Thus, $57 \leq \mu \leq 92$.

All three of these conditions must be met, so

$$82 < \mu < 86.$$

(b) Note that the first and last intervals in the list do not overlap, so they cannot both be correct. Thus, the three correct intervals Nature refers to are either: the first three or the last three. If the first three:

$$57 \leq \mu \leq 82;$$

if the last three:

$$86 \leq \mu \leq 92.$$

So the answer is that μ is in the interval $[57, 82]$ or the interval $[86, 92]$.

13. (a) Nine lower bounds are larger than 300, making nine intervals too large. One upper bound is smaller than 300, making one interval too small. Thus, $18 - 9 - 1 = 8$ intervals are correct.

(b) This means that exactly two of the lower bounds are too large. The two are 393 and 409. Thus, $\mu < 393$. But the 'exactly' tells us that the lower bound 375 is *not* too large, so $\mu \geq 375$. Putting these together, we get

$$375 \leq \mu < 393.$$

(c) This means that at least five of the upper bounds are too small. Thus, 351 is too small; the answer is

$$\mu > 351.$$

(d) By trial and error, the largest possible number of correct intervals is 11. This will happen if:

$$322 \leq \mu \leq 326 \text{ or } 344 \leq \mu \leq 345.$$

(Yes, this is a difficult question.)

14. (a) Thirteen lower bounds are larger than 600, making 13 intervals too large. Six upper bounds are smaller than 600, making six intervals too small. Thus, $24 - 13 - 6 = 5$ intervals are correct.

(b) This means that exactly three of the upper bounds are too small. The three are 518, 535 and 565. Thus, $\mu > 565$. But the 'exactly' tells us that the upper bound 568 is *not* too small, so $\mu \leq 568$. Putting these together, we get

$$565 < \mu \leq 568.$$

(c) This means that at least four of the lower bounds are too large. Thus, 720 is too large; the answer is

$$\mu < 720.$$

(d) By trial and error, the largest possible number of correct intervals is 11. This will happen if: $649 \leq \mu \leq 655$. (Yes, this is a difficult question.)