Math 202 Spring 2018 Solutions
Final Exam

Part I. **(15 points)** Consider the following data set

~~45~~ ~~35~~ ~~62~~ ~~58~~ ~~33~~,
~~42~~ ~~24~~ ~~36~~ ~~43~~ ~~72~~,
~~69~~ ~~52~~ ~~44~~ ~~38~~ ~~41~~,
~~42~~ ~~48~~ ~~37~~ ~~62~~ ~~35~~

24, 33, 35, 35, 36, 37,
38, 41, 42, 42, 43, 44, 45
48, 52, 55, 62, 62, 69, 72

1. (6 points) Compute the median, the lower and upper quartiles.

$$\text{Median} = \frac{42+43}{2} = 42.5$$

Lower quartile = 36 (5-th)
Upper quartile = 52 (15-th)

2. (4 points) Find the mean and the standard deviation.

$$\bar{x} = \frac{915}{20} = 45.75$$

$$SD = \sqrt{\frac{(45-45.75)^2 + \ldots + (35-45.75)^2}{20}} = 12.34$$

3. (2 points) What percent of the data is within one standard deviation of the mean?

# of data entries between $[45.75 - 12.34, 45.75 + 12.34]$ $= [33.41, 58.09]$ is 14;

$$\frac{14}{20} = 70\% \text{ of the data}$$

4. (3 points) Is the histogram of the data likely to be symmetric or right-skewed or left-skewed? Explain in view of your answers to 1. and 2. above.

Right-skewed since Median < Mean

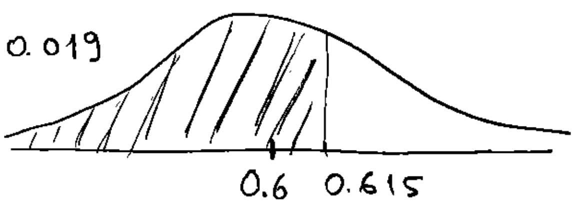
Part II. (25 points) Suppose that on the average, 60% of the graduating seniors at a certain university have at least one parent attend the graduation ceremony.

1. (8 points) In a random sample of 6 graduating seniors, what is the approximate probability that at most 4 of them will have at least one parent present in the ceremony?

$$P(X \le 4) = P(X=0) + P(X=1) + \dots + P(X=4) =$$
$$1 - P(X=6) - P(X=5) = 1 - (0.60)^6 - 6(0.60)^5(0.40) =$$
$$= 1 - 0.0467 - 0.187 = 0.767$$

2. (7 points) In a random sample of 650 graduating seniors, what is the approximate probability that at most 400 of them will have at least one parent present in the ceremony?

$$p = 0.60$$
$$SE = \sqrt{\frac{(0.60)(0.40)}{650}} = \sqrt{0.00037} = 0.019$$



$$\bar{p} = \frac{400}{650} = 0.615$$

$$z = \frac{0.615 - 0.6}{0.019} = 0.78$$

z	Table
0.78	57.63%

$$\frac{57.63\%}{2} + 50\% = 78.8\%$$

3. (10 points) In a random sample of 650 graduating seniors, it is observed that exactly 397 students had at least one parent attend the ceremony. Conduct a test of hypothesis to test the claim that more than 60% of the graduating seniors have at least one parent attend the graduation ceremony in general, at $\alpha = 1\%$.

$H_0: p = 0.60$
$H_a: p > 0.60$

z	Area
0.52	38.29

$$P = \frac{100\% - 38.29\%}{2} = 30.85\% > 1\%$$

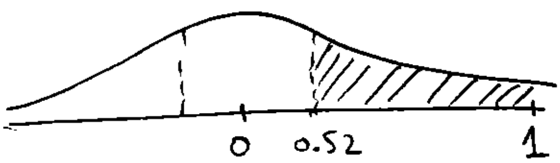
$$\bar{p} = \frac{397}{650} = 0.61$$

We don't reject the null hypothesis.

$SE = 0.019$

There are not more than 60% with at least 1 parent at the ceremony.

$$z = \frac{0.61 - 0.60}{0.019} = 0.52$$



Part III. (15 points) Two different types of chemical solution are evaluated for polishing the contact lenses to be used in the human eye. 300 lenses were polished using the first solution, and of those, 253 were satisfactory with no defects. Another 300 lenses were polished using the second solution, and 196 lenses were satisfactory in the end.

Is there any reason to believe that the two polishing solutions differ in performance?

$H_0: p_1 = p_2$

$H_a: p_1 \neq p_2$

$N_1 = N_2 = 300$

$\bar{p}_1 = \frac{253}{300} = 0.843$; $\bar{p}_2 = \frac{196}{300} = 0.653$

$SE = \sqrt{\frac{(0.843)(0.157)}{300} + \frac{(0.653)(0.347)}{300}} =$

$= \sqrt{0.00044 + 0.00076} = \sqrt{0.0012} =$

$= 0.035$

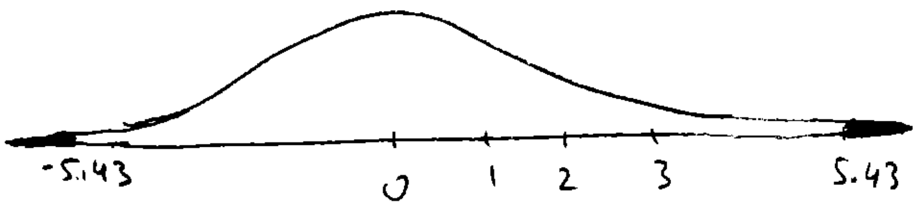
$z = \frac{0.843 - 0.653}{0.035} = 5.43$

z is very high (not on the table)

=> P is almost 0. (very low)

We reject the null hypothesis.

There is a reason to believe that the two solutions differ in performance.



Part IV. (15 points) Consider the following data set which indicates the number of traffic accidents involving children within two kilometres of schools in a city, listed by days of the week.

$100 \cdot \frac{1}{5} = 20$

Day	Number of accidents	Expected # of Accidents
Monday	23	20
Tuesday	18	20
Wednesday	17	20
Thursday	19	20
Friday	23	20
Total	100	

1. (11 points) Test the hypothesis that the number of traffic accidents occurs uniformly throughout the week; that is, on average, there are an equal number of traffic accidents each day.

$H_0: p_1 = p_2 = p_3 = p_4 = p_5 = 0.20$
H_a: at least one p is different

$$\chi^2 = \frac{(23-20)^2}{20} + \frac{(18-20)^2}{20} + \dots + \frac{(23-20)^2}{20} = \frac{1}{20}(9+4+9+1+9) = \frac{1}{20} \cdot 32 = \boxed{1.6}$$

deg. of freedom $= 5-1=4$

$1.06 < 1.6 < 2.20$

$90\% > p > 70\%$

$p > 5\%$, therefore we don't reject H_0.
On average there is an equal # of car accidents each day.

2. (4 points) The data suggest that there are a higher number of accidents on the first (Monday) and the last (Friday) day of the week.

a) Please suggest a variable related to human or social behavior, which could explain this observation. Is it qualitative or quantitative?

Level of attention, quantitative

(At the first day and last day of the week, people could be careless)

b) Also, state the possible values of the variable that you defined in a).

It can be measured on a scale of 1, ... 7 for example: Possible values: 1, ..., 7.

(Or, qualitative variable, with values: careful and careless)

Part V. **(15 points)** Is marital status related to education level? Consider the following joint frequency table to conduct an appropriate hypothesis test to answer this question.

	Middle School or Lower	High School	College or higher	
Never married	18	36	30	84
Married	12	36	80	128
Divorced	9	18	15	42
	39	90	125	254

deg. of freed = $(3-1)\cdot(3-1) = 4$

			Total
$\frac{1}{254}\cdot 84\cdot 39 = 12.8$	29.8	$\frac{1}{254}84\cdot 125 = 41.3$	84
19.7	45.3	63	128
$\frac{1}{254}\cdot 42\cdot 39 = 6.4$	14.9	$\frac{1}{254}\cdot 42\cdot 125 = 20.7$	42
Total 39	90	125	

Ho: marital status and education level are indep.

$17 = 11.3 + 5.7$

Ha: —"— —"— —— are not indep.

$$\chi^2 = \frac{(18-12.9)^2}{12.9} + \frac{(36-29.8)^2}{29.8} + \frac{(30-41.3)^2}{41.3} + \dots + \frac{(15-20.7)^2}{20.7} =$$

$= 19.18 > 13.28$ P-value $< 1\%$. We reject Ho.

There is dependence between marital status and educational level. They are correlated.

Part VI. (25 points) Samples of pottery from the Roman era as found in two different regions in Anatolia are studied for their aluminum oxide content. The first sample consisted of 5 observations and had an average aluminum oxide content of $\bar{x} = 17.3$ and variance $s^2 = 2.5$. The second sample contained 6 observations, and had a mean of 18.2 and variance $s^2 = 4.2$.

1. (10 points) Is there sufficient evidence to believe that the mean aluminum oxide content is lower in the first region?

$N_1 = 5$, $\bar{X}_1 = 17.3$, $s_1^2 = 2.5$
$N_2 = 6$, $\bar{X}_2 = 18.2$, $s_2^2 = 4.2$

$H_0: \mu_1 = \mu_2$
$H_a: \mu_1 < \mu_2$

we use t-test
deg. of freed = 9

$$SE = \sqrt{\frac{2.5}{5} + \frac{4.2}{6}} = \sqrt{1.2} = 1.09$$

$$t = \frac{(17.3 - 18.2) - 0}{1.09} = \frac{-0.9}{1.09} = -0.83$$

$0.70 < 0.83 < 1.38$
$25\% > p > 5\%$

we dont reject H_0.
There isn't sufficient evidence to believe that the 1st mean is lower.

2. (15 points) A new random sample of Roman pottery from a recent archeological excavation from a third region in Anatolia reported of $\bar{x} = 16.9$ and variance $s^2 = 5.1$ for 4 observations. Is there a significant difference between the three regions in terms of aluminum oxide content?

$H_0: \mu_1 = \mu_2 = \mu_3$
H_a: at least 1 is different.

D.f. D $= 5+6+4-3 = 12$
D.f. N $= 3-1 = 2$

$\boxed{F^*_{2,12} = 3.89}$

$$\bar{X}_{GM} = \frac{5 \cdot 17.3 + 6 \cdot 18.2 + 4 \cdot 16.9}{6+4+5} = \frac{263.3}{15} = 17.55$$

$$S_B^2 = \frac{5(17.3-17.55)^2 + 6(18.2-17.55)^2 + 4(16.9-17.55)^2}{3-1} = \frac{4.537}{2} = 2.269$$

$$S_W^2 = \frac{2.5(5-1) + 4.2(6-1) + 5.1(4-1)}{6+5+4-3} = \frac{46.3}{12} = 3.86$$

$$\frac{S_B^2}{S_W^2} = \frac{2.269}{3.86} = 0.58 < 3.89 \rightarrow$$

We don't reject H_0.
There is no significant difference between the three regions.