Math 282 FINAL EXAM Spring 2018

SHOW ALL WORK *Keep three decimal places in most calculations*

1. [8pts] Find the **mean** and **standard deviation** of the sample below.

-1 5 2 8 3 -4 0 7

$$\text{mean} = \frac{-1+5+2+8+3+(-4)+0+7}{8} = \frac{20}{8} = \boxed{2.5}$$

$$s^2 = \frac{1}{7}\left[(-1-2.5)^2+(5-2.5)^2+(2-2.5)^2+(8-2.5)^2+(3-2.5)^2+(-4-2.5)^2+(7-2.5)^2 + (0-2.5)^2\right]$$

$$= \frac{1}{7}(12.25+6.25+0.25+30.25+0.25+42.25+20.25+6.25)$$

$$= \frac{1}{7}(118.25) = 16.892857 \qquad \boxed{s = 4.11}$$

2. [16pts] The commute time (in minutes) of 17 people are as follows:

5 10 8 12 21 40 20 11 14 15 25 4 16 18 27 10 12

4 5 8 10 10 11 12 12 14 15 16 18 20 21 25 27 40

1) Find the five-number summary.

min = 4 $\quad Q_1 = 10 \quad Q_2 = 14 \quad Q_3 = 20.5 \quad$ max = 40

2) Is there any outlier in this data set? If any, which value? Use 1.5 IQR rule.

$IQR = Q_3 - Q_1 = 20.5 - 10 = 10.5 \quad 1.5\,IQR = 15.75$

upper fence $= Q_3 + 1.5\,IQR = 20.5 + 15.75 = 36.25$

Lower fence $= Q_1 - 1.5\,IQR = 10 - 36.25 = -26.25$

$\boxed{\text{40 is an outlier}}$

3) Make a stemplot for the data.

Stem	leaves
0	4 5 8
1	0 0 1 2 2 4 5 6 8
2	0 1 5 7
3	
4	0

3. [16pts] A radar unit is used to measure speeds of cars on a highway. The speeds, x, are normally distributed with a mean of 56 mph (mile per hour) and a standard deviation of 6.5 mph. The speed limit of this highway is 55 mph.

   1) What is the probability that a car picked at random is travelling at more than 65mph? X – speeds $X \sim N(56, 6.5)$

$$P(X>65) = P\left(Z > \frac{65-56}{6.5}\right) = P(Z > 1.38)$$

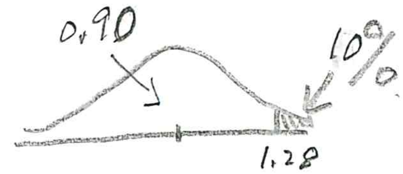
$$= 1 - 0.9162 = \boxed{0.0838}$$

   2) What is the probability that a car picked at random is travelling between 50mph and 60mph?

$$P(50 < X < 60) = P\left(\frac{50-56}{6.5} < Z < \frac{60-56}{6.5}\right) = P(-0.92 < Z < 0.6$$

$$= 0.7324 - 0.1788 = 0.5536.$$

   3) How fast must a car's speed be in order to be on the top ~~5~~10%?



$$\frac{X-56}{6.5} = 1.28$$

$$X = 1.28(6.5) + 56 = \boxed{64.32}$$

At least 64.32 mph to be on the top 10%.

   4) Take an SRS of 16 cars travelling on this highway. What is the probability that the average speed is lower than 55 mph? $\bar{X} \sim N\left(56, \frac{6.5}{\sqrt{16}}\right)$.

$$P(\bar{x} < 55)$$

$$= P\left(Z < \frac{55-56}{6.5/\sqrt{16}}\right) = P(Z < -0.62)$$

$$= 0.2676$$

4. [~~10~~ 12pts] John's parents recorded his height at various ages up to 66 months.

The data below is the summary of the records:

The correlation between John's age and height is r = 0.994.

variable	Mean	standard deviation
Age (months), x	$\bar{x}$ =52.8	s_x =11.541
Height (inches), y	$\bar{y}$ =40.4	s_y =3.975

1) Find the **intercept** and **slope** of the least square regression line.

slope $b = r \frac{s_y}{s_x} = 0.994 \frac{3.975}{11.541} = \boxed{0.3424}$

intercept $a = \bar{y} - b\bar{x} = 40.4 - 0.3424(52.8) = \boxed{22.321}$

2) Write the equation of the regression line.

$\hat{y} = 22.321 + 0.3424x$

3) Use the equation you found in 2) to predict John's height when he was 50 months.

$22.321 + 0.3424(50) = 39.441$ (inches)

4) John's parents decide to use the least-squares regression line to predict his height at age 21 years (252 months). We conclude (circle one)

a) John's parents will get a fairly accurate estimate of his height at age 21 years because the data are clearly correlated.

b) Such a prediction could be misleading because it involves extrapolation.

5. [10pts] A couple plans to have three children. There are 8 possible arrangement of girls and boys.

a) Write down all 8 arrangements of the sexes of three children.

bbb, bgb, gbb, bbg, ggb, gbg, bgg, ggg

b) Let X be the number of boys the couple will have. Find the distribution of X.

X	0	1	2	3
P	1/8	3/8	3/8	1/8

6. [15pts] The data below summarize " Helmet Use and Risk of Head Injuries in Alpine Skiers and Snowboarders"

	Head Injuries	Not Injured	total
Wore Helmet	96	656	752
No Helmet	480	2330	2810
total	576	2986	3562

If one of the subjects is randomly selected,

a) Find the probability of selecting someone with a head injury.

$$\frac{576}{3562} = 0.1617$$

b) Find the condition probability of selecting someone with a head injury given the subject wore helmet.

$$\frac{96}{752} = 0.1277$$

c) Are "wore helmet" and "head injury" independent? Why?

No, They are not independent, because P(head injury) ≠ P(head injury / wore helmet).

7. [12pts] A multiple choice test has eight questions, each with four answer choices. One of these choice is the correct answer. Assume you make random guesses. Let X be the number of questions you guessed correctly, then X has a binomial distribution with $n = 8$, $p = 0.25$.

a) Find the probability that you guessed exact ~~4~~ 3 questions correctly.

$$\binom{8}{3} 0.25^3 (1-0.25)^5 = 56\,(0.015625)(0.2373)$$
$$= 0.2076$$

b) Find the probability that you guessed at least one question correctly.

$$1 - \binom{8}{0} 0.25^0 (1-0.25)^8 = 1 - 0.75^8 = 1 - 0.10011$$
$$= 0.9$$

8. [~~10~~ 10pts] Suppose there are 100 multiple choice questions in question 7, then x has a binomial distribution with $n = 100$, $p = 0.25$.

a) Find the mean and standard deviation of X.

mean $= np = 100(0.25) = \boxed{25}$ standard deviation $= \sqrt{np(1-p)}$
$= \sqrt{25(0.75)} = \boxed{4.33}$

b) Use the normal approximation to find the probability P(X<30).

$$P(X<30) = P\left(Z < \frac{30-25}{4.33}\right) = P(Z<1.15)$$
$$= 0.8749$$

9. [~~15~~ 16 pts] The distribution of times (in number of days) until maturity of a certain variety of tomato plants is Normally with mean μ and standard deviation σ = 2.4. An SRS of 9 plants of this variety yields an average time of maturity $\bar{x}$ =65.5 days. Answer the questions below.

   1) To make a confidence interval for the mean mature time μ, which procedure do you use?
      a) Z-procedure   b) t-procedure

   2) Make a 95% confidence interval for the mean mature time μ.

      $\bar{x} \pm z^* \frac{\sigma}{\sqrt{n}} = 65.5 \pm 1.96 \frac{2.4}{\sqrt{9}} = 65.5 \pm 1.568$

      95% CI : (63.932, 67.068)

   3) Interpret the confidence interval you found in 2).

      We are 95% confident that mean mature time of this variety tomato plants is between 63.932 to 67.068 days.

   4) If you want the margin of error for the 95% confidence interval to be ± 1 day, you should select a simple random sample of size
      a) 7   b) 39   c) 23   d) 22

      $n = \left(\frac{z^* \sigma}{m}\right)^2 = \left(\frac{1.96(2.4)}{1}\right)^2 = 22.128$

10. [10pts] Diet colas use artificial sweeteners to avoid sugar. **You want to test that these sweeteners gradually lose their sweetness over time**. Trained tasters sip the cola and score the cola on a "sweetness score" of 1 to 10. The cola then stored for a month at high temperature to imitate the effect of four months' storage at room temperature. **This is a matched – pairs design.** Our data are the difference (score before storage – score after storage) in the tasters' scores. A size of 10 simple random sample yields a sample mean $\bar{x}$ = 1.02 and standard deviation s = 0.9. Suppose the difference of the scores has a Normal distribution with mean μ. Do an appropriate four steps hypotheses test.

    1) $H_0 : \mu = 0$ $H_a : \mu > 0$.

    2). $t = \frac{\bar{x}}{s/\sqrt{n}} = \frac{1.02}{0.9/\sqrt{10}} = 3.5839$. $df = 9$

    3) $0.0025 < \text{P-Value} < 0.005$

    4). Reject H_0: Strong evidence that sweeteners gradually lose their sweetness over time.

11. [12pts] A national health survey of 1472 U.S. adults (selected randomly) during 2010 revealed that 677 had never smoked cigarettes. $\hat{p} = \frac{677}{1472} = 0.4599$

1) Make a 90% confidence interval for *p*, the proportion of U.S. adults in 2010 that have never smoked.

$$\hat{p} \pm Z^* \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = 0.4599 \pm 1.645 \sqrt{\frac{0.4599(1-0.4599)}{1472}}$$

$$0.4599 \pm 1.645(0.01299) \Rightarrow 0.4599 \pm 0.0214$$

$$\Rightarrow 90\%\ CI: (0.4385, 0.4813) \text{ or } (43.85\%, 48.13\%)$$

2) In 1965, about 44% of the U.S. adults had never smoked cigarettes. You suspect that *p*, the proportion of U.S. adults that have never smoked has grown since 1965 and want to use the sample from 2010 to test: $H_0: p = 44\%$, $H_a: p > 44\%$ or $H_0: p = 0.44$, $H_a: p > 0.44$

a) Calculate the test statistics.

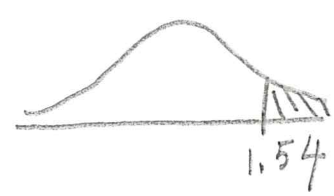
$$Z = \frac{\hat{p} - P_0}{\sqrt{\frac{P_0(1-P_0)}{n}}} = \frac{0.4599 - 0.44}{\sqrt{\frac{0.44(1-0.44)}{1472}}}$$

$$= \frac{0.0199}{0.01294} = 1.54$$

b) Find the p-value.

$$\text{P-Value} = 1 - 0.9382$$

$$= 0.0618$$



c) Write your conclusion at α=0.05 significant level.

Fail to reject H_0. There is no enough evidence that the proportion of U.S. adults that have never smoked has grown since 1965.

12. [~~10~~ 12 pts] Is there a **difference** in the amount of airborne bacteria between carpeted and uncarpeted room? In an experiment, 9 rooms were carpeted and 8 were uncarpeted. The rooms are similar in size and function. After a suitable period of time, the concentration of bacteria in the air was measured (in unites of bacteria per cubic foot) in all of these rooms. The summarized data are provided:

	Sample size	average concentration	sample standard deviation
Carpeted rooms	$n_1 = 9$	$\bar{x}_1 = 184$	$s_1 = 22.0$
Uncarpeted rooms	$n_2 = 8$	$\bar{x}_2 = 175$	$s_2 = 16.9$

Let μ_1 = mean concentration of bacteria in the air of all the similar carpeted rooms.
μ_2 = mean concentration of bacteria in the air of all the similar uncarpeted rooms.
Do an appropriated 4 steps hypotheses test.

1) $H_0: \mu_1 = \mu_2$ $\quad H_a: \mu_1 \neq \mu_2$

2) $$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} = \frac{184-175}{\sqrt{\frac{22^2}{9} + \frac{16.9^2}{8}}} = \frac{9}{9.4593}$$

$$= 0.95$$

3) $0.3 < \text{P-Value} < 0.4$ $\quad df = 8-1 = 7$

4) Do not reject H_0. There is no enough evidence to conclude that there is a difference in the amount of airborne bacteria bacteria between carpeted and uncarpeted rooms.

13. [18pts] It is desired to test whether there was any relationship between education level and smoking status. The two-way table below summarizes the data from a simple random sample of 459 men.

Education level		Smoking Status				total
		Nonsmoking	Former	Moderate	Heavy	
Primary school	Observed	56	54	41	36	187
	Expected	(59.48)	(50.926)	(42.37)	(34.222)	
	Cell chisq	[0.204]	[0.186]	[0.044]	[0.092]	
Secondary school	Observed	37	43	27	32	139
	Expected	(44.214)	(37.854)	(31.495)	(25.438)	
	Cell chisq	[1.177]	[0.67]	[0.642]	[1.693]	
University	Observed	53	28	36	16	133
	Expected	(42.305)	(36.22)	(30.135)	(24.34)	
	Cell chisq	[2.704]	[1.866]	[1.141]	[2.858]	
Total		146	125	104	84	459

a) Find the value of the two expected counts and the 2 cell chi square contribution that need to be computed. Show work.

$$Exp_{3\times1} = \frac{146\times133}{459} = 42.305$$

$$Cell\ chisq_{3\times1} = \frac{(53-42.305)^2}{42.305} = 2.704$$

$$Exp_{3\times4} = \frac{84\times133}{459} = 24.34$$

$$Cell\ chisq_{3\times4} = \frac{(16-24.34)^2}{24.34} = 2.858$$

b) Do a 4 step of hypotheses. Show how the appropriate table is used (α = 0.05).

1) H_0: No relation between education level and smoking sta

H_a: Some relation between education level and smoking stat

2) $\chi^2 = 0.204 + 0.186 + 0.044 + 0.092 + 1.177 + 0.67 + 0.642 + 1.693 + 2.704 + 1.866 + 1.141 + 2.858 = 13.277$

$df = (3-1)\times(4-1) = 2\times3 = 6$

3) $0.025 < P\text{-value} < 0.05$

[33pts] 4). Reject H_0. There is evidence that no relation betwe education level and smoking status.

14. Circle the correct answer(s).

1) If A, B are two disjoint events, then A and B are independent.
   a) True b) false

2) If the correlation between two variables is close to 0, you can conclude that a scatterplot would show
   a) A cloud of points with no visible pattern. b) A strong straight-line pattern.
   c) No straight-line pattern, but there might be a strong pattern of another form.

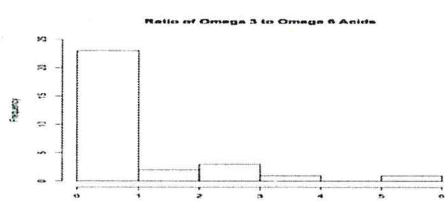
3) A political party sends a mail survey to 1500 randomly selected registered voters in a community. The survey asks respondents to give an opinion about the job performance of the current President. Of the 1500 surveys sent out, 480 are returned, and of these, only 120 say they're satisfied with the President's job performance. The sample is
   a) The 1500 randomly selected voters receiving the questionnaire.
   b) The 120 voters surveyed that are satisfied with the President's job performance.
   c) The voters in his district. d) The 480 surveys returned.

4) To assess the opinion of students at SIU about campus safety, you went to Morris Library late at a night interviewed 20 students who are willing to give their opinion. The method of sampling used is
   a) Simple random sampling. b) Voluntary response sampling.
   c) Stratified random sampling. d) Convenience sampling

5) Street ice cream sales and the rates of violent crime and murder are positive correlated. So we may conclude that buying ice cream on street causes crime.
   a) True b) false. There are might be lurking variables.

6) A, B are two independent events, P(A) = 0.5, P(B) = 0.3. The condition probability P(A/B) is:
   a) 0.5 b) 0.3 c) 0.15 d) 1.67 e) 0.6

7) The following is a histogram of a data set. The histogram is



   a) Skewed to left. b) Skewed to right c) Has one possible high outlier

8) Which numerical summary would you choose for the data in question 7)?
   a) Five-number summary b) mean and standard deviation

9) Based on an SRS of 300 houses from a city, the 95% confidence interval of the average price in the city is ($120,000, $180,000). You random selected a house, then there is a 95% chance that the price of the house is between $120,000 and $180,000.
   a) True b) False

10) When constructing a confidence interval for the mean of a normal distribution in the case when σ is known, the length of the interval is widest when the confidence level and sample size are: level=(99/90/80) and sample size=(80/60/40)
   a) level=80, sample size=40 b) level=99, sample size=40 c) level=99, sample size=80

11) How large a sample size n would you need to estimate population proportion *p* with margin of error 0.03 ~~with~~ and 95% confidence? Assume that you don't know anything about the value of *p*.
   a) 16 b) 17 c) 1068 d) 1067.11

$$n=\left(\frac{z^*}{m}\right)^2 p^*(1-p^*)$$

$$p^*=0.5$$

$$n=\left(\frac{1.96}{0.03}\right)^2 0.5(1-0.5)=1067.1$$