Final Exam Practice Problems- Math 263-009 - Fall 13 -Kennedy

1. An effort is currently under way to cap (or limit) awards given by the courts to patients who successfully sue doctors for malpractice. If the cap were established, and if the cap affected only a few very large awards, check all true statements:

   FALSE The median award would be reduced

   TRUE The mean award would be reduced

   FALSE The IQR (inter-quartile range) of the awards would be reduced

   TRUE The standard deviation of the awards would be reduced.

   TRUE The range of the awards would be reduced.

2. Malaria infection rates may be affected by changing climates. To test this, data was collected from four cities in west Africa: in each city, the total daily precipitation (in mm) was measured, and the percent of the citys population who tested positive for malaria was measured.

rainfall (mm)	2.1	3.2	4.8	7.8
percentage	4.5	4.2	3.8	3.0

   (a) Name the explanatory variable and give its units.

   (b) Name the response variable and give its units.

   (c) Find the regression line and the value of the correlation coefficient r. **Solution:**

$$y = -0.2622x + 5.049, \quad r = -0.99992$$

   (d) Give the units of the y-intercept, and then interpret the value of the y-intercept in the context of this problem. **Solution:** Units of y-intercept are the units of the response variable which has no units.

   (e) Give the units of the slope, and then interpret the value of the slope in the context of this problem. **Solution:** The units of the slope are the units of the response variable divided by the units of explanatory variable which gives 1/mm. The interpretation of the slope is that for every 1 mm reduction in averages daily rainfall the percentage of malaria cases decreases on average by 0.26%.

(f) What can you conclude from the value of r? **Solution:** There is a negative association of rainfall and malaria - the more it rains the lower the malaria rate on average. Since $|r|$ is so close to 1, the association is very strong.

3. A group of doctors has a new drug that is supposed to reduce the incidence of heart attacks. They give the drug to patients who have a family history of heart attacks. After a year they look at how many patients who got the drug had a heart attack and how many who did not get the drug had a heart attack.

   (a) What is wrong with this experimental design? **Solution:** The patients are not randomly assigned to the drug/no-drug groups. The doctors might be prescribing the drug more often for patients they think are at higher risk of heart attack in which case the drug/no-drug groups would not be comparable.

   (b) How would you change the experimental design to fix this problem? **Solution:** A third party should randomly assign the patients to two groups. One group gets the drug, the other group gets an identical looking placebo. The doctors don't know which group each patient is in and the patients don't know (double blind).

4. In a certain large city, 72% of the people own a cell phone, 38% have a land line, and 29% have both a cell phone and a land line.

   (a) What proportion of people in the city own neither a cell phone nor a land line? **Solution:** The proportion that own either a cell or landline or both is $72\% + 38\% - 29\% = 81\%$. So 19% own neither.

   (b) If we choose a person who owns a cell phone at random, what is the probability that this person also has a land line? **Solution:** $0.29/0.72 = 0.403 = 40.3\%$.

   (c) If we choose a person who has a land line at random, what is the probability that this person also owns a cell phone? **Solution:** $0.29/0.38 = 0.763 = 76.3\%$.

(d) Are having a land line and owning a cell phone independent events? Explain why or why not.

$P(cell \cap landline) = 0.29, \quad P(cell)P(landline) = 0.72 \times 0.38 = 0.273$

Since $0.29 \neq 0.273$ they are not independent.

5. The probability distribution of a random variable X is given in the table.

value	8	12	15	20
probability	0.2	0.1	0.4	0.3

(a) Compute the mean and variance of X. **Solution:**

$$\begin{aligned}\mu &= 0.2 \times 8 + 0.1 \times 12 + 0.4 \times 15 + 0.3 \times 20 = 14.8 \\ \sigma^2 &= 0.2 \times (8 - 14.8)^2 + 0.1 \times (12 - 14.8)^2 \\ &\quad +0.4 \times (15 - 14.8)^2 + 0.3 \times (20 - 14.8)^2 = 18.16 \\ \sigma &= 4.26\end{aligned}$$

(b) Find the probability that X is greater than 14. **Solution:** $0.4 + 0.3 = 0.7$.

(c) Find the probability that X is greater than 14 given that X is even. **Solution:**

$$P(X > 14|Xeven) = \frac{P(X > 14 \cap even)}{P(Xeven)} = \frac{0.3}{0.2 + 0.1 + 0.3} = 0.5$$

6. Male diastolic blood pressure (in mmHg) is normally distributed, with mean 72 and standard deviation 6.

(a) A large group of men each measures their diastolic blood pressure. What percent of them will find a blood pressure above 80? **Solution** Calculator: $normcdf(80, 10000, 72, 6) = 0.0912$

(b) A random sample of 3 men each measures their diastolic blood pressure. They then take the average of these 3 measurements. What is the probability that they will find an average blood pressure above 80? **Solution** So the sample of three the standard deviation is $6/\sqrt{3} = 3.464$ Calculator: $normcdf(80, 10000, 72, 3.464) = 0.0105$

(c) A certain drug is aimed at men who have blood pressure higher than 99% of all men. What blood pressure measurement will these men have? **Solution** Calculator: $invnorm(0.99, 72, 6) =$ 85.96mmHg.

7. A newspaper surveys the inhabitants of a city to ask whether they are satisfied with the number of hospitals in the city.

   (a) The newspaper reports that 30%±4% are satisfied, with 95% confidence. Assuming that the newspaper did its statistics correctly, how many people did the newspaper survey? **Solution** The margin of error is $z^*\sqrt{\hat{p}(1-\hat{p})/n}$. For 95% confidence $z^* = 1.96$. So $0.04 = 1.96\sqrt{0.3(1-0.3)/n}$. Solving for n gives $n = 505$.

   (b) The newspaper editor is unsatisfied with the ±4% and the 95% confidence. In order to reduce the 4% to 1% and increase the 95% to 99%, how many people must the newspaper survey? **Solution** For 99% confidence we need $z^* = 2.57$. So $0.01 = 2.57\sqrt{0.3(1-0.3)/n}$. Solving for n gives $n = 13,870$.

8. An ANOVA test (or F-test) is performed to compare IQ scores in Arizona, New Mexico, Texas, and Nevada.

   (a) State the null hypothesis and the alternate hypothesis. Define any parameters you use.

   Solution: H_0 : The means of IQ in the four states are all equal. H_a : At least two of the four means are different.

   (b) A computer performs the F-test and finds a p-value of 0.98. What conclusion can you make about IQ scores in Arizona, New Mexico, Texas, and Nevada?

   Solution: This is not a small probability. In particular it is not less than any reasonable significance level. So we do not reject H_0. Based on the data we cannot conclude the mean IQ scores differ for any two of the states.

In the next four problems you should

- State the null hypothesis and the alternate hypothesis.
- Calculate the statistic.

- Find the P-value.
- State your conclusion about the null hypothesis.
- State your conclusion in terms of the original question.

9. On March 15, 2005, a court in Oakland, California, was told that for 25 years a judge had attempted to exclude Jews from juries in death row cases because they would not vote for the death penalty. Court records show that during this time, 29 people who were Jewish or had Jewish- sounding names had been called to serve on a jury for a death row case. Of these, 27 were excluded. Is there evidence that Jews were excluded more often than the rest of the population, whose exclusion rate is 49.97%? Use a significance level of $\alpha = 5\%$.

   Solution: Let p be the proportion of prospective jurors who were Jewish or had Jewish sounding names who are excluded. $H_0 : p = 0.4997$. $H_a : p > 0.4997$. We have $\hat{p} = 27/29 = 0.931$.

$$z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} = \frac{0.931 - 0.4997}{\sqrt{\frac{0.4997(1-0.4997)}{29}}} = 4.645$$

   This gives a p-value of $normcdf(4.645, 10000, 0, 1) = 0.0000017$ which is way below the significance level. So we reject the null hypothesis. This data indicates that jurors who are Jewish or have Jewish sounding names are more likely to be excused than the rest of the population.

10. A 1999 study surveyed 1500 college students (600 male, 900 female) nationwide about their drinking. Among other questions, students were asked whether they had drank alcohol in the past month. 438 men reported having had an alcoholic drink in the past month, while 603 women reported having had an alcoholic drink in the past month. Decide whether college students of one gender were more likely to drink than the other gender. Use a significance level of $\alpha = 2\%$. **Solution:** Let p_1 be the proportion of male students who have had a drink in last 6 months, p_2 be the proportion of female students who have had a drink in last 6 months. $H_0 : p_1 = p_2$. $H_a : p_1 \neq p_2$. We have $\hat{p}_1 = 438/600 = 0.73$, $\hat{p}_2 = 603/900 = 0.67$, To compute the z statistic we use a pooled estimator for $p_1 = p_2$: $\hat{p} = (438 + 603)/(600 + 900) = 0.694$. Then

$$z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}(1-\hat{p})(\frac{1}{n_1} + \frac{1}{n_2})}} = 2.47$$

The p-value is $2 \times 0.0067 = 0.0135$ which is less than the given significance level of 0.02. So we reject the null hypothesis and conclude that one gender is more likely to have had a drink in the past six months.

11. Six students were chosen at random to test two new shoe types. Each student wore one type of shoe and ran a 100 meter sprint; the next day, each student wore the other type of shoe and ran a 100 meter sprint. Decide whether students run significantly faster in one type of shoe than the other. Use a significance level of $\alpha = 2\%$. **Solution:** For each student we compute the difference of his or her time in shoe B minus his or her time in shoe A: The population is the difference

	Student 1	Student 2	Student 3	Student 4	Student 5	Student 6
Shoe A	15.5	18.3	14.5	11.3	12.0	13.7
Shoe B	15.5	19.9	13.5	11.7	12.8	14.9
B-A	0.0	1.6	-1.0	0.4	0.8	1.2

in a student's time in shoe B and shoe A. Let μ be the mean of this difference. $H_0 : \mu = 0$. $H_a : \mu \neq 0$. For this sample of 6 difference we compute the sample mean and variance: $\overline{x} = 0.5$, $s = 0.927$. Then we use a t statistic for one population:

$$t = \frac{\overline{x} - \mu_0}{s/\sqrt{n}} = \frac{0.5 - 0}{0.927/\sqrt{6}} = 1.32$$

This has a t-distribution with 5 degrees of freedom. The p-value is $2 \times 0.122 = 0.244$. This is well above the significance level, so we do not reject the null hypothesis. This data does not provide evidence that one shoe helps you run faster than the other.

12. A researcher studied the relationship between coffee drinking and heart attacks in young women. The table below summarizes the results.

    (a) Find the probability that a randomly selected light coffee drinker has a heart attack. **Solution:** $88/(172 + 88) = 0.338$

    (b) Are the amount of coffee drunk and heart attacks related? Use a significance level of $\alpha = 5\%$. **Solution:** H_0 : coffee drinking and chances of a heart attack are not related. H_a : coffee drinking

observed	No attack	attack	
light drinker	172	88	260
moderate drinker	132	70	202
heavy drinker	45	40	85
	349	198	

and chances of a heart attack are related. We use the χ^2 statistic. First compute the row and column sums:
We have $N = 547$. The expected counts are :

expected	No attack	attack
light drinker	165.9	94.1
moderate drinker	128.9	73.1
heavy drinker	54.2	30.8

We have

$$\chi^2 = \frac{(165.9-172)^2}{165.9} + \frac{(94.1-88)^2}{94.1} + \frac{(128.9-132)^2}{128.9} + \frac{(73.1-70)^2}{73.1} + \frac{(54.2-45)^2}{54.2} + \frac{(30.8-40)^2}{30.8} = 5.17$$

The number of degrees of freedom is $(2-1)(3-1) = 2$. This gives a p-value of 0.0754. This is greater than 0.05, so we do not reject the null hypothesis. This evidence does not show a relation between coffee drinking and heart attackes at the 5% significance level.

13. The nation-wide average salary of a full professor is \$84,173 with standard deviation \$22,063. The mean salary for 15 randomly chose full professors at the U of A was \$79,143. Assuming the standard deviation for the U of A population is the same as the national population, find a 85% confidence interval for the mean salary of a full professor at the U of A.

    Solution: Since the standard deviation is known, the CI is $\overline{x} \pm z^* \sigma / \sqrt{n}$. For the 85% confidence level we want probability of 0.075 in each tail. $invnorm(0.075, 0, 1) = -1.4395$, so $z^* = 1.4395$. So the CI is

$$\$79,143 \pm 1.44 \times \$22,063/\sqrt{15} = \$79,143 \pm \$8,200$$

So the interval is $[\$70943, \$87343]$

14. Here is a list of the statistics we have seen:

    (a)
    $$z = \frac{\overline{x} - \mu_0}{\sigma/\sqrt{n}}$$

    (b)
    $$z = \frac{\overline{x}_1 - \overline{x}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

    (c)
    $$t = \frac{\overline{x} - \mu_0}{s/\sqrt{n}}$$

    (d)
    $$t = \frac{\overline{x}_1 - \overline{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

    (e)
    $$z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$$

    (f)
    $$z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}(1-\hat{p})(\frac{1}{n_1} + \frac{1}{n_2})}}$$

    (g) χ^2 (chi-squared)

    (h) ANOVA (or F)

    For each of the following situations, choose an appropriate statistic for the hypothesis test. Some of them may have two correct answers; in this case, you should give both correct answers.

- **(f)** Decide which of two tuberculosis drugs is more likely to work. (Each drug either cures a person or fails to cure a person.)
- **(d)** Decide which of two tuberculosis drugs works more quickly.
- **(f) or (g)** Decide whether the gender of a job applicant affects whether the applicant is offered a job.
- **(g) (assuming there are three or more ethnicities)** Decide whether the ethnicity of a job applicant affects whether the applicant is offered a job.
- **(d) or (h)** Decide whether the gender of a job applicant affects the applicants starting salary.
- **(h)** Decide whether the ethnicity of a job applicant affects the applicants starting salary.
- **(e)** Decide which of two candidates for a school election is likely to win.
- **(c)** Decide whether a certain city's average annual income is above the national average. **Comment:** this problem is vague. The intent was to assume that we knew the national average and only sample people in the city. If you interpret the question as calling for sampling both the city and the nation, then the answer would be (d).