NAME: Key **Math 214 – Exam 3**

Justify answers and show all work for full credit.

Problem 1.

An SRS of 85 students is taken from a large university to estimate the proportion of students whose parents bought a car for them. In the sample, the parents of 51 students bought them a car.

(a) What is a 95% confidence interval for the population proportion p of students whose parents bought them a car?

(b) Suppose you want to test whether more than half of all students have parents buy them cars.

4 State the hypotheses H_0 : $p = \frac{1}{2}$ H_A: $p > \frac{1}{2}$

(c) Specify the statistic and compute the value of the statistic.

(d) Estimate the P-value for this hypothesis test.

(e) At the the 5% confidence level, what is the conclusion? What does it mean in this case?

a) $\hat{p} = \frac{51}{85} = 0.6$ $\hat{p} \pm z^* \sqrt{\hat{p}(1-\hat{p})/n}$

6 $0.6 \pm 1.96\sqrt{(0.6)(0.4)/85}$

$\underline{0.6 \pm 0.104}$ ie. $(0.496, 0.704)$

b) $z = \frac{\hat{p} - p}{\sqrt{p(1-p)/n}}$

4 $= \frac{(0.6-0.5)}{\sqrt{(0.5)(0.5)/85}} = \frac{0.1}{0.0542} = 1.84$

4 d) $P(z \geq 1.84) = 0.0329$

4 e) $0.0329 < \alpha = 0.05$, So we reject H_0

There is strong evidence that more than half of all students at this university have parents buy them cars.

110 22

Problem 2.

Are parents of girls more inclined to buy them a car than parents of boys? An SRS of 40 girls and 45 boys is taken from a large university. In the samples, 20 girls and 21 boys responded that their parents bought them a car. Use girls as Group 1 and boys as Group 2.

(a) What is an estimate of the <u>difference</u> between the true proportions for boys and girls?

(b) Suppose you want to test whether or not the true proportions are the same.

4 State the hypotheses H_0 : $p_1 = p_2$ $\qquad H_A$: $p_1 \neq p_2$

(c) Specify the statistic and compute the value of the statistic.

(d) Estimate the P-value for this hypothesis test.

(e) At the the 5% confidence level, what is the conclusion? What does it mean in this case?

a) $\hat{p}_1 = \frac{20}{40} = 0.500 \quad \hat{p}_2 = \frac{21}{45} = 0.467$

4 $D = \hat{p}_1 - \hat{p}_2 = 0.033$

c) Use pooled $SE_D = \sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$, $\quad \hat{p} = \frac{20+21}{85} = 0.482$

$= \sqrt{(0.482)(0.518)\left(\frac{1}{40} + \frac{1}{45}\right)}$

$= 0.1086$

6 $z = \frac{\hat{p}_1 - \hat{p}_2}{SE_D} = \frac{0.5 - 0.467}{0.1086} = 0.304$

4 d) $P(z \geq 0.304) \approx 0.3821 \qquad 2P = 0.764$

6 e) $2P > \alpha = 0.05$, So we <u>fail to reject H_0</u> ← 2pts.

2 There is <u>not</u> enough evidence to conclude that $p_1 \neq p_2$.

24

Problem 3.

100 college students were asked what size pizza they order and what is their favorite topping.

Size	Pepperoni	Topping: Veggie	Cheese	Total
Small	18	11	6	35
Medium	14	12	7	33
Large	3	9	20	32
Total	35	32	33	100

(a) State the null hypothesis H_0 for a chi-square test based on this data.

(b) Assuming H_0, what is the expected count for a medium pepperoni pizza?

(c) What is the contribution to the chi-square statistic for the medium pepperoni pizza cell?

(d) What are the degrees of freedom df for the chi-square statistic?

(e) Suppose the chi-square statistic for this data is 11.5. What is the P-value?

(f) At the the 5% confidence level, what is the conclusion? What does it mean in this case?

4 a) There is no association between rows and columns

4 b) $(33)(35)/100 = 11.55$

4 c) $(14-11.55)^2/11.55 = 0.520$

4 d) $(3-1)(3-1) = 4$

4 e) $0.02 < P < 0.025$

4 f) $P < \alpha = 0.05$ So we reject Ho

2 There is evidence that size and toppings are associated.

26

Problem 4. SHORT ANSWER and MULIPLE CHOICE

To study the relationship between sales and profits, an SRS of 79 companies from the Forbes 500 list was used, and the following results were obtained from statistical software:

$R^2 = 0.662$ and $s = 466.2$

Variable	Parameter estimate	Standard error
Constant	–176.644	61.16
Sales	0.092498	0.0075

3 **(1)** What is the value of the intercept of the least-squares regression line? –176.644

5 **(2)** What is the 90% confidence interval for the slope b_1? $0.0925 \pm$ 0.0125

$b_1 \pm t^* SE_{b_1} = 0.0925 \pm (1.664)(0.0075)$

4 **(3)** Suppose the researchers conducting the study wish to test the hypotheses H_0: $b_1 = 0$ versus H_a: $b_1 > 0$. What do we know about the P-value of this test?

$t = \frac{b_1}{SE_{b_1}} = 12.3$

A) The P-value is greater than 0.10.

B) The P-value is between 0. 05 and 0.10.

C) The P-value is between 0.01 and 0.05.

D) The P-value is less than 0.01.

E) There is not enough information to determine the P-value range.

4 **(4)** Is there strong evidence of a straight-line relationship between sales and profits?

A) Yes, because the slope of the least-squares line is positive.

B) Yes, because the P-value for testing if the slope is 0 is quite small.

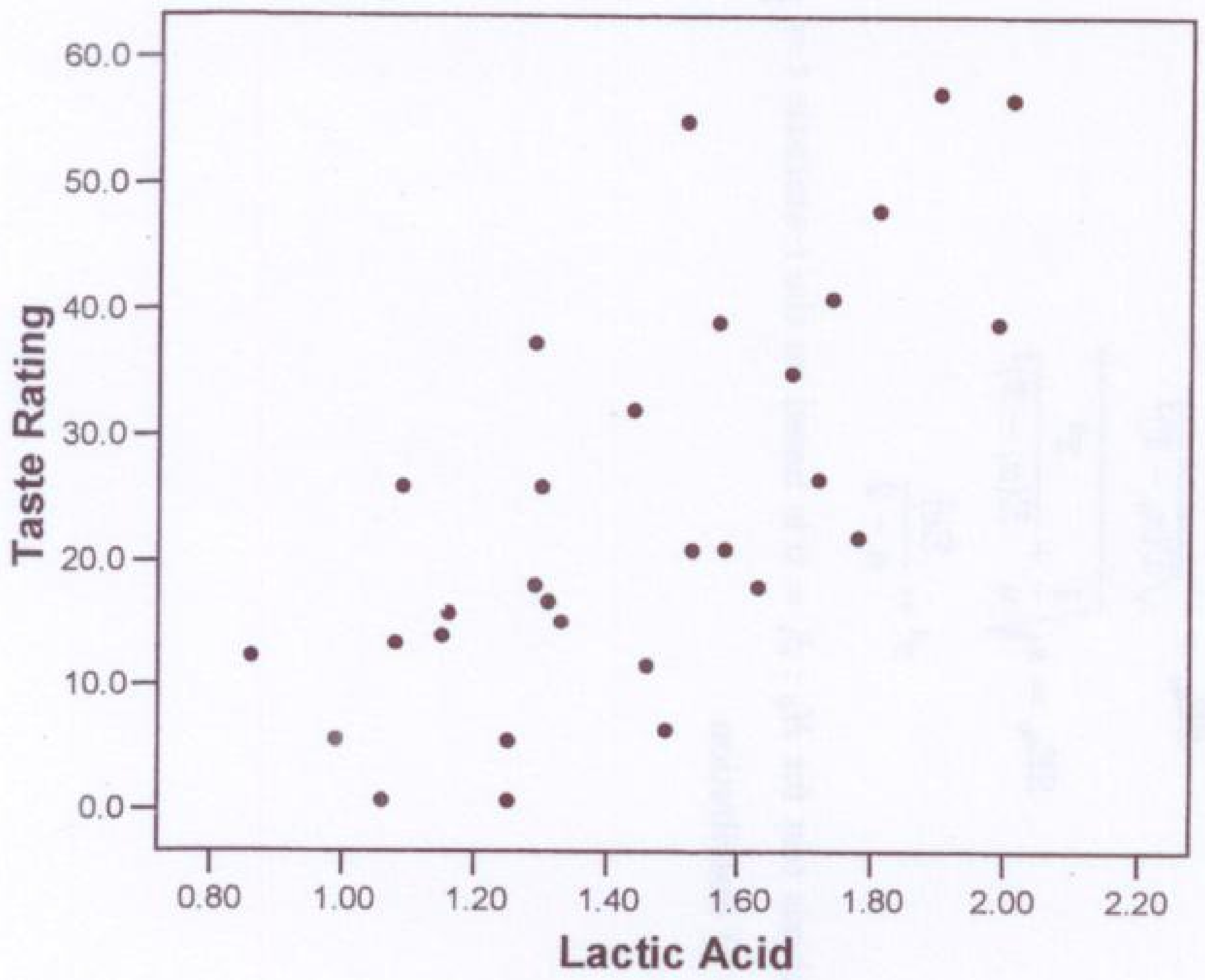
C) No, because the P-value for testing if the slope is 0 is quite big.

D) No, because the value of the square of the correlation is relatively small.

E) It is impossible to say, because we are not given the actual value of the correlation.

16

Problem 5. MULIPLE CHOICE and TRUE/FALSE



4

(1) What is a plausible value for the correlation between lactic acid and taste rating?

A) 0.9 B) 0.7 C) 0.07 D) –0.07 E) –0.7 F) –0.9

(2) Determine whether each of the following statements is True or False.

8

A) F The least-squares regression line is the line that makes the square of the correlation in the data as large as possible.

B) T The least-squares regression line is the line that makes the sum of the squares of the vertical distances of the data points from the line as small as possible.

C) F The least-squares regression line is the line that best splits the data in half, with half of the points above the line and half below the line.

D) T The least-squares regression line always passes through the point ($\bar{x}$, $\bar{y}$), the means of the explanatory and response variables, respectively.

(3) Researchers studied a sample of 300 adults and found a strong negative correlation between the amount of vitamin X an individual consumed and the number of pounds he/she was overweight.

8

Determine whether each of the following statements is True or False.

A) F This is quite strong evidence that large amounts of vitamin X cause adults to not gain weight.

B) T If the amount of vitamin X consumed and the number of pounds overweight for each individual in this study were plotted on a scatterplot, the points would lie close to a negatively sloping straight line.

C) F If a larger sample of adults had been studied, the correlation would have been even stronger.

D) T If one of the individual data points had a negative residual, then it lies below the regression line.

20

Formulas

The statistic $\bar{x}$ has

$$\mu_{\bar{x}} = \mu, \quad \sigma_{\bar{x}} = \sigma/\sqrt{n}$$

and if the population is normal, or n is large enough, is approximately normally distributed.

The statistic $D = \bar{x}_1 - \bar{x}_2$ has

$$\mu_D = \mu_1 - \mu_2$$

If the two random samples are independent of each other then $\sigma_D^2 = (\sigma_1^2/n_1 + \sigma_2^2/n_2)$. The distribution of D will be approximately normal.

We have discussed symmetric confidence intervals given by these formulas

$$\bar{x} \pm z^* \sigma/\sqrt{n}, \quad \bar{x} \pm t^* s/\sqrt{n}, \quad (\bar{x}_1 - \bar{x}_2) \pm t^* SE, \quad \hat{p} \pm z^* \sqrt{\hat{p}(1-\hat{p})/n}$$

where the unspecified value SE may be one of two values depending on an assumption about equality of the standard deviations in the samples.

We discussed using the following test statistics of the form observed minus expected divided by SE: $z = \dfrac{\bar{x} - \mu}{\sigma}\sqrt{n}$,

$$t = \frac{\bar{x} - \mu}{s}\sqrt{n}, \quad t = \frac{\bar{x}_1 - \bar{x}_2}{s_p\sqrt{1/n_1 + 1/n_2}}, \quad t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{s_1^2/n_1 + s_2^2/n_2}}, z = \frac{\hat{p} - p}{\sqrt{p(1-p)/n}}$$

For those with a t-distribution, you need to remember the degrees of freedom associated with each.

For testing differences in two sample proportions against the null hypothesis: $p_1 = p_2$ (pooled)

$$D = \hat{p}_1 - \hat{p}_2, \quad z = \frac{\hat{p}_1 - \hat{p}_2}{SE_D}, \quad SE_D = \sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)} \quad \hat{p} = \frac{X_1 + X_2}{n_1 + n_2}$$

χ^2 Statistic:

$$\chi^2 = \sum \frac{(\text{Observed} - \text{Expected})^2}{\text{Expected}}$$

Means and Standard Deviations:

$$\bar{x} = \frac{1}{n}\Sigma x_i \qquad s^2 = \frac{1}{n-1}\Sigma(x_i - \bar{x})^2$$

Correlation Coeficient and Regression Lines:

$$r = \frac{1}{n-2}\Sigma\frac{(x_i - \bar{x})}{s_x}\frac{(y_i - \bar{y})}{s_y}$$

$$\hat{y} = b_1 x + b_0$$

where:

$$b_1 = r\frac{s_y}{s_x}$$
$$b_0 = \bar{y} - b_1\bar{x}$$

$$e_i = y_i - \hat{y}_i$$

Confidence Intervals for population regression parameters β_0 and β_1:

$$b_1 \pm t^*\mathrm{SE}_{b_1} \qquad b_0 \pm t^*\mathrm{SE}_{b_0}$$

$$\mathrm{SE}_{b_1} = \frac{s}{\sqrt{\Sigma(x_i - \bar{x})^2}}$$

$$\mathrm{SE}_{b_0} = s\sqrt{\frac{1}{n} + \frac{\bar{x}^2}{\Sigma(x_i - \bar{x})^2}}$$

$$s^2 = \frac{\Sigma e_i^2}{n-2}$$

The hypothesis test for $H_0 : \beta_1 = 0$ is based on the t-statistic $t = \dfrac{b_1}{\mathrm{SE}_{b_1}}$ and the $t(n-2)$ distribution.