Math 171 **Final Exam** (Sample)

Two formula sheets allowed.

1) A SRS of twenty-five male faculty members at a large university found that ten felt that the university was supportive of female and minority faculty. An independent SRS of twenty female faculty found that five felt that the university was supportive of female and minority faculty. Let p_1 and p_2 represent the proportion of all male and female faculty, respectively, at the university who felt that the university was supportive of female and minority faculty at the time of the survey.

a) (5%) Is there evidence that the proportion of male faculty members who felt the university was supportive of female and minority faculty is larger than the corresponding proportion for female faculty members? To make this determination, test the hypothesis

$$H_0 : p_1 = p_2, \qquad H_a : p_1 > p_2$$

(at the $\alpha = 0.05$ significance level)

b) (5%) Find a 95% confidence interval for $p_1 - p_2$.

2) All current-carrying wires produce electromagnetic (EM) radiation, including the electrical wiring running into, through, and out of our homes. High frequency EM radiation is thought to be a cause of cancer; the lower frequencies associated with household current are generally assumed to be harmless. To investigate this possibility, researchers visited the addresses of children in the Denver area who had died of some form of cancer (leukemia, lymphoma, or some other type) and classified the wiring configuration outside the building as either a high-current configuration (HCC) or a low-current configuration (LCC). Here are some of the results of the study.

	Leukemia	Lymphoma	Other cancers
HCC	52	10	17
LCC	84	21	31

The Minitab output for the above table is given below. The output includes the cell counts, the expected cell counts, and the chi-square statistic. Expected counts are printed below observed counts.

	Leukemia	Lymphoma	Other cancers	Total
HCC	52 49.97	10 11.39	17 17.64	79
LCC	84 86.03	21 19.61	31 *	136
Total	136	31	48	215

ChiSq $= 0.082 + 0.170 + 0.023 + 0.048 + 0.099 + * = **$

a) (5%) Find the expected count of LCC/other cancers and the chisquare value of the test.

b) (5%) Is there strong evidence of an association between wiring configuration and the type of cancer from which the children in the study died? Write the conclusion using the langugage of the context.

3) Salary data of 1992-1993 for a sample of fifteen universities was obtained. We are curious about the relation between mean salaries for assistant professors (junior faculty, x, say) and full professors (senior faculty, Y, say) at a given university. Suppose the true regression line is

$$EY = \mu_Y = \alpha + \beta x$$

and that the assumptions for regression inference were satisfied. This model was fit to the data using least squares. The following results were obtained from statistical software. Note that salaries were in thousands of dollars. Assistant professor salaries were treated as the explanatory variable x and full professor salaries as the response variable Y.

Predictor	**Coefficient**	**St.Dev.**
Constant	15.0658	14.36
x	1.40827	0.3217

$S = 5.503$ R-Sq= 0.596, $\overline{x} = 44$, $\sum(x_i - \bar{x})^2 = 292.6$.

a) (5%) Find the least squares regression equation according to the data.

b) (5%) Find a 90% confidence interval for β.

c) (5%) Is there strong evidence that straightline dependence on x has value for predicting Y ? Specifically, test at the 10% significance level $H_0 : \beta = 0$ vs $H_1 : \beta \neq 0$.

d) (5%) Suppose a particular university has $x = 40$. Give a 90% prediction interval for the corresponding value of Y for this university.

4) Charles, Julia and Alex are in grades 4, 3, and 2, respectively, and are representing their school at a spelling bee. The school's team score is the sum of the number of words the individual students spell correctly out of 50 words each. Different words are given for each grade level. From practicing at school, it is known that the probability of spelling each word correctly is 0.9 for Charles and 0.8 for the younger two, Julia and Alex. Assume the individual scores are independent.

a) (6%) Find the mean and standard deviation for the number of correct words for each child.

b) (6%) Find the mean and standard deviation for the team score.

c) (5%) Does the team score have a binomial distribution? Explain.

d) (6%) Assume that the team score is approximately normal. If last year's team score was 131, what is the approximate probability that this year's team scores as well or better?

5) Suppose a one-sample test t-test of $H_0 : \mu = 0$ versus $H_a : \mu \neq 0$ results in a statistic of $t = 0.65$ and $df = 14$. Suppose a new study is done with $n = 150$, and the sample mean and standard deviation turn out to be exactly the same as in the first study.

a) (6%) What conclusion would you make in the original study, using $\alpha = 0.05$?

b) (6%) What conclusion would you make in the new study, using $\alpha = 0.05$?

6) Assume that the American female college students' heights follow a normal distribution with mean 65.5 inches. Similarly, the American male college students' heights have a normal distribution with mean 67.5 inches. Both of the populations have a standard deviation of 2.5 inches.

a) (6%) What is the probability that a randomly chosen male college student is taller than 5 feet 10 inches (70 inches)?

b) (6%) Randomly choose a male college student and a female college student. What is the probability that he is taller than her?

c) (6%) Randomly pick 5 male students. What is the probability that the average height of these 5 students is higher than 70 inches?

d) (7%) Randomly pick 5 male students and 4 female students. What is the probability that the males' average is higher than the females' average?

Solutions

Problem 1) **a)** We use a two sample test for proportions. The sample proportion for males is $\hat{p}_1 = 10/25 = 0.4$ with sample size $n_1 = 25$. The sample proportion for females is $\hat{p}_2 = 5/20 = 0.25$ with sample size $n_2 = 20$. The pooled sample proportion is

$$\hat{p} = \frac{n_1\hat{p}_1 + n_2\hat{p}_2}{n_1 + n_2} = \frac{10+5}{45} = 0.33.$$

The test statistic is

$$\begin{aligned} z &= \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}} = \frac{0.4 - 0.25}{\sqrt{(0.33)\cdot(1-0.33)\left(\frac{1}{25}+\frac{1}{20}\right)}} \\ &= 1.06. \end{aligned}$$

The P-value is $P(Z > 1.06) = 0.1446$, so H_0 is not rejected. The critical value is z^* is the upper 5% quantile of the standard normal, i.e. $z^* = 1.65$.

b) Note first that the confidence interval is two-sided, so the quantile z^* to be used here is different from that in the one-sided test in a). The confidence interval is

$$\hat{p}_1 - \hat{p}_2 \pm z^* \cdot SE$$

where z^* is the upper 2.5% quantile of the standard normal, i.e.

$$P(Z > z^*) = 0.025,\ z^* = 1.96$$

and SE is the standard error (i.e. estimated standard deviation) of $\hat{p}_1 - \hat{p}_2$:

$$\begin{aligned} SE &= \sqrt{\hat{p}_1(1-\hat{p}_1)/n_1 + \hat{p}_2(1-\hat{p}_2)/n_2} \\ &= \sqrt{(0.4)(1-0.4)/25 + (0.25)(1-0.25)/20} \\ &= 0.14. \end{aligned}$$

The interval is thus

$$\begin{aligned} & 0.4 - 0.25 \pm 1.96 \cdot 0.14 \\ = \ & 0.15 \pm 0.27 \\ = \ & (-0.12, 0.42). \end{aligned}$$

Problem 2) **a)** The expected count for the cell LCC/other cancer is

$$\text{expected count} = \frac{\text{row total} \times \text{column total}}{\text{table total}} = \frac{136 \cdot 48}{215} = 30.36.$$

The chisquare value for this cell then is

$$\frac{(\text{observed count-expected count})^2}{\text{expected count}} = \frac{(31 - 30.36)^2}{30.36} = 0.0135$$

and the chisquare value for the test

$$\begin{aligned} & 0.082 + 0.170 + 0.023 + 0.048 + 0.099 + 0.0135 \\ = \ & 0.436 \end{aligned}$$

b) Set significance level $\alpha = 0.05$. For testing the null hypothesis of independence, the critical value (for the chisquare distribution with $(2-1)(3-1) = 2$ degrees of freedom) is $\chi^2_* = 5.99$ and the P-value is $P = P(\chi^2 > 0.436) > 0.25$, so the null hypothesis of independence cannot be rejected. There is no significant evidence of an association between wiring configuration and the type of cancer.

Problem 3) a) The equation is

$$\hat{y} = a + bx$$

where $a = 15.07$ is the estimated intercept, $b = 1.41$ is the estimated slope, thus

$$\hat{y} = 15.07 + 1.41 \cdot x.$$

b) The confidence interval is

$$b \pm t^* \cdot SE_b$$

where SE_b is given in the software output as $SE_b = 0.32$ and $P(t > t^*) = 0.05$ for a t with $15 - 2 = 13$ degrees of freedom. Thus $t^* = 1.77$ and the interval is

$$\begin{aligned} & 1.41 \pm (1.77) \cdot (0.32) \\ = \; & 1.41 \pm 0.57 \\ = \; & (0.84, 1.98). \end{aligned}$$

c) The answer is given by the confidence interval in b): since 0 is outside this interval, the null hypothesis $H_0 : \beta = 0$ is rejected against the two-sided alternative. The confidence level 90% matches the significance level of the test (10%).

Alternatively, if the test is formally carried out (anew), then the test statistic is

$$t = \frac{b}{SE_b} = \frac{1.41}{0.32} = 4.41$$

and the two sided P-value for 13 degrees of freedom is $P = 2 \cdot P(t > 4.41) \approx 2 \cdot (0.0005) = 0.001$. Thus $P < 0.1$ and the null hypothesis is rejected.

d) The predicted value of $\hat{y}$ is

$$\hat{y} = a + b \cdot 40 = 15.07 + 1.41 \cdot 40 = 71.47.$$

The prediction interval is

$$\hat{y} \pm t^* \cdot SE_{\hat{y}}$$

where for $x^* = 40$

$$SE_{\hat{y}} = s \cdot \sqrt{1 + \frac{1}{n} + \frac{(x^* - \bar{x})^2}{\sum (x_i - \bar{x})^2}}.$$

We obtain

$$\begin{aligned} SE_{\hat{y}} &= (5.503) \cdot \sqrt{1 + \frac{1}{15} + \frac{(40 - 44)^2}{292.62}} \\ &= 5.83. \end{aligned}$$

Hence the interval is (t^* can be taken from b))

$$\begin{aligned} & 71.47 \pm 1.77 \cdot 5.83 \\ = \; & 71.47 \pm 10.32 \\ = \; & (61.15, 81.79)\,. \end{aligned}$$

Problem 4) **a)** The number of correct words for Charles (X_C, say) is a random variable with a binomial distribution $B(n,p)$ with $n = 50$ and probability of success $p = 0.9$. Thus

$$\begin{aligned} EX_C &= np = 50 \cdot (0.9) = 45 \\ SD(X_C) &= \sqrt{np(1-p)} = \sqrt{45 \cdot (0.1)} = 2.12. \end{aligned}$$

Similarly the number of correct words for Julia (X_J, say) and Alex (X_A, say) is $B(n,p)$ with $n = 50$ and $p = 0.8$, thus

$$\begin{aligned} EX_J &= np = 50 \cdot (0.8) = 40 \\ SD(X_J) &= \sqrt{40 \cdot (0.2)} = 2.83. \end{aligned}$$

b) The team score is

$$X_T = X_C + X_J + X_A$$

thus

$$EX_T = EX_C + EX_J + EX_A = 45 + 2 \cdot 40 = 125.$$

The variance of the team score is (since X_C, X_J, X_A are independent)

$$\begin{aligned} \mathrm{Var}\,(X_T) &= \mathrm{Var}\,(X_C + X_J + X_A) = \mathrm{Var}\,(X_C) + \mathrm{Var}\,(X_J) + \mathrm{Var}\,(X_A) \\ &= 4.5 + 2 \cdot 8 = 20.5, \\ SD(X_T) &= \sqrt{20.5} = 4.53 \end{aligned}$$

c) No, since the sucess probabilities are different (0.9 and 0.8). The sum $X_C + X_J + X_A$ is the number of successes in 150 independent trials, but these trials do not have all equal sucess probability.

d) We assume that X_T is approximately normal; the mean and standard deviation were calculated in b). The probability is

$$\begin{aligned} P\,(X_T > 131) &= P\left(\frac{X_T - EX_T}{SD(X_T)} > \frac{131 - EX_T}{SD(X_T)}\right) \\ &\approx P\left(Z > \frac{131 - 125}{4.53}\right) = P\,(Z > 1.32) \\ &= 0.0934 \end{aligned}$$

Problem 5) **a)** The critical value for rejecting H_0 against the two-sided alternative at $\alpha = 0.05$, for $df = 14$ is $t^* = 2.145$, thus H_0 is not rejected. The P-value is

$$\begin{aligned} P &= 2 \cdot P(t > 0.65) = 2 \cdot P(t > 0.65) > 2 \cdot 0.25, \text{ thus} \\ P &> 0.5. \end{aligned}$$

b) The t-statistic for the one sample problem is

$$t = \frac{\bar{x}}{s/\sqrt{n}}.$$

We are given the value of this statistic (t_{ol}, say) in the original study $t_{ol} = 0.65$ for $n = df + 1 = 15$. Let t_{ne} be the value of the t-statistic for the new study with $n = 150$, and $\bar{x}$ and s are the same as in the original study. Since

$$t_{ol} = \frac{\bar{x}}{s/\sqrt{15}}, \; t_{ne} = \frac{\bar{x}}{s/\sqrt{150}}$$

we have the relationship

$$t_{ne} = t_{ol} \cdot \sqrt{10}.$$

We thus obtain

$$t_{ne} = 0.65 \cdot \sqrt{10} = 2.06$$

The critical value for rejecting H_0 against the two-sided alternative at $\alpha = 0.05$, for $df = 100$ is $t^* = 1.984$, for $df = 1000$ it is $t^* = 1.962$. For $df = 149$ the critical value is between those two, thus the test rejects.

Problem 6) **a)** Let X_F be the random variable representing a randomly chosen female student, $\mu_F = EX_F = 65.5$ and σ be the standard deviation $\sigma = 2.5$. Similarly, let X_M be the random variable representing a randomly chosen male student, $\mu_M = EX_M = 67.5$. Then

$$\begin{aligned} P(X_M > 70) &= P\left(\frac{X_M - \mu_M}{\sigma} > \frac{70 - \mu_M}{\sigma}\right) \\ &= P\left(Z > \frac{70 - 67.5}{2.5}\right) = P(Z > 1) \\ &= 0.1587. \end{aligned}$$

b) First compute mean and standard deviation of the difference $X_M - X_F$. We have

$$\begin{aligned} E(X_M - X_F) &= \mu_M - \mu_F = 2 \\ \text{Var}(X_M - X_F) &= \text{Var}(X_M) + \text{Var}(X_F) \\ &= 2 \cdot \sigma^2, \\ SD(X_M - X_F) &= \sqrt{2} \cdot \sigma = \\ &= \sqrt{2} \cdot 2.5 = 3.54. \end{aligned}$$

The probability is

$$\begin{aligned} P(X_M > X_F) &= P(X_M - X_F > 0) \\ &= P\left(\frac{X_M - X_F - 2}{3.54} > \frac{0-2}{3.54}\right) \\ &= P(Z > -0.56) = P(Z < 0.56) \\ &= 0.7123. \end{aligned}$$

c) Let $\bar{X}_M$ be the average of the randomly chosen 5 male students. Then

$$\begin{aligned} E\bar{X}_M &= EX_M = 67.5, \\ SD(\bar{X}_M) &= SD(X_M)/\sqrt{5} = \sigma/\sqrt{5}. \end{aligned}$$

Thus

$$\begin{aligned} P(\bar{X}_M > 70) &= P\left(\frac{\bar{X}_M - \mu_M}{\sigma/\sqrt{5}} > \frac{70-\mu_M}{\sigma/\sqrt{5}}\right) \\ &= P\left(Z > \sqrt{5}\right) = P(Z > 2.24) \\ &= 0.0125. \end{aligned}$$

d) Let $\bar{X}_F$ be the average of the randomly chosen 4 female students. Then

$$\begin{aligned} E\bar{X}_F &= EX_F = 65.5, \\ \mathrm{Var}(\bar{X}_F) &= \mathrm{Var}(X_F)/4 = \sigma^2/4 \end{aligned}$$

Thus we have

$$\begin{aligned} E(\bar{X}_M - \bar{X}_F) &= \mu_M - \mu_F = 2 \\ \mathrm{Var}(\bar{X}_M - \bar{X}_F) &= \mathrm{Var}(\bar{X}_M) + \mathrm{Var}(\bar{X}_F) \\ &= \sigma^2/5 + \sigma^2/4 = \sigma^2 \cdot \left(\frac{1}{5} + \frac{1}{4}\right) \end{aligned}$$

$$\begin{aligned} SD(\bar{X}_M - \bar{X}_F) &= \sigma \cdot \sqrt{\left(\frac{1}{5} + \frac{1}{4}\right)} = 2.5 \cdot 0.67 \\ &= 1.68. \end{aligned}$$

Consequently

$$\begin{aligned} P(\bar{X}_M - \bar{X}_F > 0) &= P\left(Z > -\frac{2}{1.68}\right) \\ &= P(Z > -1.19) = P(Z < 1.19) \\ &= 0.883. \end{aligned}$$