Math 171

Practice Final

1. (20%) There is some evidence that the attitude of cancer patients can influence the progress of their disease. We cannot experiment with humans, but some research has been done on rats. Consider the following experiment.

60 rats are injected with tumor cells, then randomly divided into two groups of 30. All the rats then receive electric shocks, but rats in Group 1 can end the shock by pressing a lever. (Rats learn this sort of thing quickly.) The rats in Group 2 cannot control the shocks, which presumably makes them feel helpless and unhappy. We suspect that the rats in Group 1 will develop fewer tumors. The results: 14 of the Group 1 rats and 19 of the Group 2 rats developed tumors.

a) (5%) State null and alternative hypotheses for this investigation.

b) (10%) Carry out the appropriate test at the 5% level of significance and state your conclusion in context.

c) (5%) Suppose further research eventually reveals that your conclusion in this study was actually incorrect. Did you commit a Type I or Type II error?

2. (15%) Suppose you were asked to analyze each of the following situations described below. (Note: You need not actually perform any of the calculations!) For each of the five parts of this problem, indicate:

i. which inference procedure you would use (from the list below):
   - one–sample mean
   - one–sample mean of differences (matched pairs)
   - difference of means, independent samples
   - one–sample proportion
   - two–sample difference of proportions
   - homogeneity or independence of categorical variables
   - regression – inference for β

ii. which statistic (z, t, or χ^2) would be involved

iii. if t or χ^2, the number of degrees of freedom

a) (3%) A child development researcher wonders if attitudes toward spanking are the same across the country. She surveys 200 randomly selected parents from each of four regions. In the Northeast, 63 approve of spanking. In the South, Midwest, and West, the number expressing approval was 105, 88 and 66 respectively.

b) (3%) Cornell University claims that even though their published costs are higher than at Ithaca College, Cornell's students receive more financial aid. A SRS of 53 Cornell students found that they received an average of \$8257 in financial aid, with a standard deviation of \$3222. At Ithaca College, the average aid for 44 randomly selected students was \$6381 with a standard deviation of \$2771.

c) (3%) A researcher wishes to know whether physical coordination and intelligence are linearly associated in young children. He samples 36 three–year–olds and gives each of them a coordination test (with possible scores from 0 to 80) and a standard IQ test.

d) (3%) In Ithaca every garbage bag must have a tag and cannot exceed a specified weight. One day a disposal crew randomly selects 135 bags and finds that 24 of these are overweight. The city wants to come up with a confidence interval for the percentage of residents who are not tagging their bags appropriately.

e) (3%) A random sample of 62 people complaining of indigestion are given an antacid. They report that their discomfort subsided in an average of 13 minutes; the standard deviation was four minutes. The manufacturer wants a 95% confidence interval for the "relief time."

3. (15%) Researchers at the University of Texas Medical Center recently selected 757 patients at random, and tested them for hepatitis C. They also found out whether these patients had tattoos, and if so, whether they had gotten the tattoos in state–licensed tattoo parlors. A Chi–square analysis of the data is shown.

	Hep-C	No H-C	Total
Tattoo (parlor)	31 7.74	62 85.26	93
Tattoo (other)	10 6.08	63 66.92	73
No tattoo	22 49.18	569 *	591
Total	63	694	757

ChiSq = 69.90 + 6.35 +
2.53 + 0.23 +
15.02 + ** = ______

a) (3%) Write appropriate hypotheses.

b) (4%) Show how to calculate the expected count (*) and the component of Chi-square (**) for the last cell.

c) (4%) Find the number of degrees of freedom, and the P–value for this test.

d) (4%) State a complete conclusion in the context of this study. (If you find significance, be sure to identify the source of the difference.)

4. (10%) Leah is flying back to Ithaca from Florida, with a connection in Pittsburgh. The probability her first flight leaves on time is 0.15. If her first flight is on time, the probability her luggage wil make the connecting flight in Pittsburgh is 0.90, but if the first flight is delayed the probability her luggage will make it is 0.65.

a) (5%) Are the first flight leaving on time and her luggage making the connection independent events? Explain briefly.

b) (5%) What is the probability that Leah's first flight is delayed and her luggage arrives in Ithaca with her anyway?

5. (15%) A man who moves to a new city sees that there are two routes he could take to work. A neighbor who has lived there a long time tells him Route A will average 5 mintes faster than Route B. The man decides to experiment. Each day he flips a coin to determine which way to go, driving each route 20 days. He finds that Route A took an average of 40 minutes with standard deviation 3 minutes, and Route B took an average of 43 minutes with standard deviation 2 minutes. It appears that travel times for the routes are approximately normally distributed.

a) (10%) Find a 95% confidence interval for the difference in average commuting time for the two routes.

b) (5%) Should the man believe the old- timer's claim that he can save an average of 5 minutes a day by always driving Route A? Explain.

6. (25%) The computer output shown provides information about the size (in square feet) of 18 Ithaca homes, and the city's assessed valuation of those homes.

a) (2%) Briefly explain why you think it is appropriate to use a linear model in this situation.

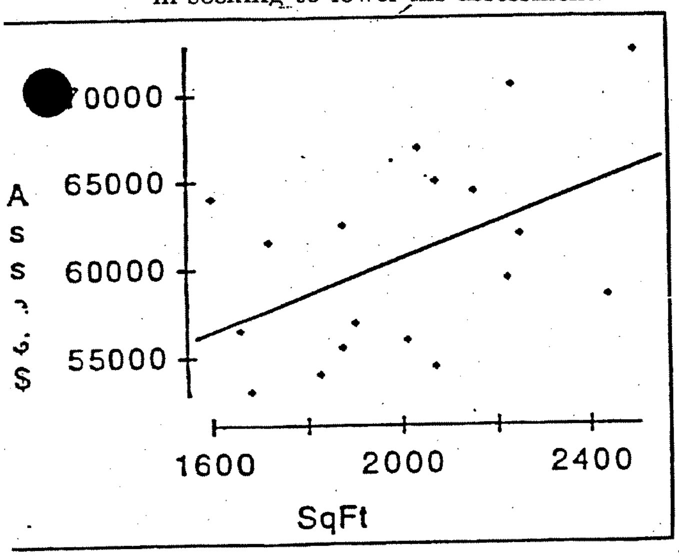
b) (3%) What is the equation of the least squares line?

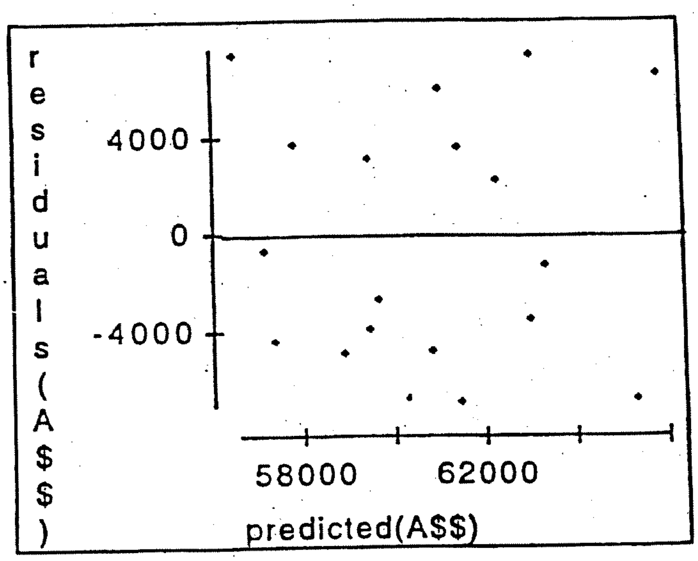
c) (2%) Explain in context the meaning of the slope of that line.

d) (2%) What percentage of the variability of assessment is explained by this model?

e) (6%) Give the P–value for testing $H_0 : \beta = 0$ against $H_\alpha : \beta \neq 0$, where β is the slope of the true regression line for all homes. Can you reject H_0 at the $\alpha = 5\%$ level of significance? State your conclusion in context.

f) (10%) The owner of a house measuring 2100 square feet claims that the $70000 asessment is too high. To make a stronger argument he hires you to construct a 95% prediction interval for the assessment of a house with 2100 square feet. Find the interval, and explain whether this provides the homeowner with a good argument in seeking to lower his assessment.





Summary of SqFt

Count	18
Mean	2005.83
StdDev	262.159

Summary of Asse$$

Count	18
Mean	60777.8
StdDev	5706.61

Dependent variable is: Asse$$

R squared = 22.8%
s = 5170 with 18 - 2 = 16 degrees of freedom

Variable	Coefficient	s.e. of Coeff	t-ratio	prob
Constant	39947.4	9670	4.13	0.0008
SqFt	10.3849	4.783	2.17	0.0453

① Two-sample proportion

Solutions

② 4a. (i) homogeneity / independence, (ii) χ^2, (iii) df = 3.

4b. (i) difference of means, independent samples, (ii) t, (iii) df = 43.

4c. (i) proportion, 1 sample, (ii) z, (iii) not applicable.

4d. (i) mean, 1 sample, (ii) t, (iii) df = 61.

---

③ C

1a. Null hypothesis H_0 is the statement that there is no relationship between having hepatitis C and having tattoos (including the type of parlor where the tattoos were gotten). The alternative is that hepatitis C and tattoos do have a relationship.

1b. Expected count $= \frac{\text{(row total)(column total)}}{\text{table total}} = \frac{(591)(694)}{757} = 541.82$, while the component of Chi-square is $\frac{(\text{observed count - expected count})^2}{\text{expected count}} = \frac{(569-541.82)^2}{541.92} = 1.36$. This makes ChiSq $= 95.39$.

1c. Degrees of freedom $= (3-1)(2-1) = 2$, and the P-value is $P(\text{ChiSq} \geq 95.39) < 0.0005$.

1d. From the test, we conclude that we have strong evidence against, H_0, or in other words that hepatitis C and tattoos are related. In particular, patients who got tattoos at a state parlor have the largest chance to get hepatitis C while patients without tattoos have the least chance to get hepatitis C.

2a. We can use a one-sample t-procedure on the differences, since we are working with matched-pair data. The 90% confidence interval is $\bar{x} \pm t^* \frac{s}{\sqrt{n}} = 3600 \pm 1.734 \frac{1071}{\sqrt{19}} = (3173.95, 4026.05)$, where $t^* = 1.734$ is the upper .05 critical value for the $t(18)$-distribution.

2b. We are 90% confident that the average extra tuition charged by a state college to a non-resident is in the interval $(3173.95, 4026.05)$.

④

3a. Let A be the event that Leah's first flight leaves on time, and let B be the event that Leah's luggage makes her connecting flight. Then $P(B|A) = 0.9$ and $P(B|A^c) = 0.65$. Thus, $P(B|A) \neq P(B|A^c)$, which means that the probability that the luggage makes the connection depends on whether or not the first flight leaves on time. Therefore, the events A and B are not independent.

3b. $P(A^c \cap B) = P(B|A^c)P(A^c) = (.65)(1-.15) = .5525$.

⑤ t - confidence interval

---

⑥

7a. The residuals do not have any apparent pattern. That is, the residual plot shows a uniform and random scatter with no curvature, no increasing or decreasing spread, and no extreme outliers or x-values.

7b. $\hat{y} = 39947.4 + 10.3849x$ (or Asse\$\$ $= 39947.4 + 10.3849$SqFt).

7c. The slope says that on the average, the assessed value of a home increases 10.3849 dollars for each increase of 1 square foot.

7d. 22.8%, the value of r^2.

7e. The data has a t-statistic of 2.17 for the hypothesis $H_0 : \beta = 0$, and P-value of 0.0453 for the two-sided alternative $H_a : \beta \neq 0$. Thus we can reject H_0 at the $\alpha = 5\%$ level, meaning that our data gives strong evidence that there is a linear relationship between assessed value and size of Ithaca homes.

7f. The prediction interval is $\hat{y} \pm t^* SE_{\hat{y}}$, where $\hat{y} = 39947.4 + 10.3849(2100) = 61755.7$, $t^* = 2.120$ (critical value of $t(16)$-distribution for a 95% confidence interval), and $SE_{\hat{y}} = \sqrt{s^2 + \frac{s^2}{n} + (x^* - \bar{x})^2 SE_b^2} = \sqrt{5170^2 + \frac{5170^2}{18} + (2100 - 2005.83)^2 4.783^2} = 5330.73$. Thus the prediction interval is $61755.7 \pm 2.120(5330.73) = (50454.6, 73056.8)$. Since the interval contains 70000, it does not give the homeowner good evidence to lower his assessment.