

**Statistics 200 Honors
Final Exam
Tuesday, December 14, 1999**

Name: _____

This 30-point test has five pages, front and back. Show as much of your work as you can; I will give partial credit where appropriate.

Work quickly; you will have only 1 hour and 50 minutes to complete as much of the test as you can.

You may use your textbook and any papers you wish.

Question 1 (2 points) Suppose that the heights of 25-year-old males are approximately normally distributed with mean 71 inches and standard deviation 2.5 inches.

(a) How tall are the tallest 2.5% of males? (1 point)

The z-value corresponding to $p = 0.975$ is 1.96. Solve for X :

$$\frac{X - 71}{2.5} = 1.96$$

This gives $X = 75.9$, so the tallest 2.5% are 75.9 inches or taller.

(b) What is the probability that a male selected at random is shorter than 68 inches? (1 point)

$$P(X < 68) = P\left(\frac{X - 71}{2.5} < \frac{68 - 71}{2.5}\right) = P(Z < -1.2) = 0.1151.$$

Question 2 (3 points) Data on 224 students who started as computer science majors at a particular university are collected. Their math SAT scores are analyzed to determine whether high school males score higher than high school females on the math SAT in the United States. Relevant Minitab output is shown below.

gender	N	Mean	StDev	SE Mean
Female	79	565.0	82.9	???
Male	145	611.8	84.0	???

95% CI for mu (Female) - mu (Male): (-69.8, -23.7)

T-Test mu (Female) = mu (Male) (vs <): T = ?????? P = 0.0000 DF = 162

(a) Find the values of SE Mean for the female and male samples. (1 point)

$$SE_F = 82.9/\sqrt{79} = 9.33 ; SE_M = 84.0/\sqrt{145} = 6.98.$$

(b) Calculate the value of the appropriate t statistic. Show your work. (1 point)

$$t = \frac{\bar{x}_1 - \bar{x}_2}{SE_{\bar{x}_1 - \bar{x}_2}} = \frac{565.0 - 611.8}{\sqrt{\frac{82.9^2}{79} + \frac{84.0^2}{145}}} = \frac{-46.8}{11.65} = -4.02$$

(c) Although the p-value for the test is very small, we may still have doubts as to whether high school males score higher than high school females on the math SAT in the United States. Explain the reason for these doubts. (1 point)

A sample of computer science students at a single university is not an SRS from the population of interest; it is almost certainly biased.

Question 3 (5 points) A man claims to have the mystic ability to influence the outcome of a fair coin flip. To test this, an experiment is performed in which he is asked to predict 125 times in succession what will happen on the flip of a fair coin. Let p be the probability that the man predicts correctly on a given flip.

(a) The null hypothesis says that the man can do no better than guessing. We are interested in knowing whether he can do better than this. In terms of p , write down the null and alternative hypotheses. (1 point)

$$H_0 : p = 0.5$$

$$H_a : p > 0.5$$

(b) Let X denote the number of times out of 125 that he is correct. Under H_0 , what are the mean and standard deviation of X ? (2 points)

X has a binomial $(125, 0.5)$ distribution. Therefore, its mean is $125(0.5) = 62.5$ and its standard deviation is $\sqrt{125(0.5)(0.5)} = 5.59$.

(c) The sample proportion in this case is equal to $\hat{p} = X/125$. If the observed value of \hat{p} is 56.8%, give the z-statistic and p-value for the test in part (a). (2 points)

$$z = \frac{\hat{p} - p_0}{\sigma_{\hat{p}}} = \frac{0.568 - 0.50}{\sqrt{\frac{0.5(0.5)}{125}}} = \frac{.068}{.045} = 1.52.$$

$$P(Z > 1.52) = 0.0643.$$

Question 4 (4 points) A recent study of college students at a large public university was carried out to determine whether a student's choice of major has anything to do with whether that student receives a student loan to pay for a portion of his or her education. Here are the results from a SRS of 583 students:

Rows: Field of Study	Columns: Student Loan		All
	no	yes	
English	58.30 137	41.70 98	100.00 235
Liberal arts and education	58.22 124	41.78 89	100.00 213
Management	68.00 51	32.00 24	100.00 75
Science	48.33 29	51.67 31	100.00 60
All	58.49 341	41.51 242	100.00 583

Chi-Square = 5.353, DF = ????, P-Value = ??????

Cell Contents --

% of Row

Observed Count

(a) Circle every one of the eight cells for which the observed count is higher than the expected count. Explain your reasoning. (2 points)

Any cell in the yes column with row percent greater than 41.51 should be circled (there are three of these). Any cell in the no column with row percent greater than 58.49 should be circled (there is one of these).

(b) How many degrees of freedom should be used here? Find the p-value for the chi-square statistic of 5.353. (2 points)

On 3 degrees of freedom, the p-value is 0.1477.

Question 5 (6 points) In the sample of 224 incoming computer science students described in Question 2, a regression model is fit using high school math (hsm), science (hss), and english (hse) grades to predict mathematics SAT (satm) scores. The relevant Minitab output is given below.

The regression equation is

$$\text{satm} = 426 + 26.0 \text{ hsm} + 1.64 \text{ hss} - 7.50 \text{ hse}$$

Predictor	Coef	StDev	T	P
Constant	426.24	32.36	13.17	0.000
hsm	26.017	3.903	6.67	0.000
hss	1.640	4.130	0.40	0.692
hse	-7.502	4.255	-1.76	0.079

$$S = 76.96 \quad R\text{-Sq} = 21.7\% \quad R\text{-Sq}(\text{adj}) = 20.7\%$$

(a) Your colleague notices the p-value of 0.692 and says he is surprised because it means that high school science grades are not significantly correlated with math SAT scores in this sample. Explain why your colleague is incorrect. (1 point)

The table above doesn't reveal what the relationship between hss and satm would be outside of this model.

(b) A friend of yours who was in this study had high school math, science, and english grades of 9, 10, and 7, respectively (on the scale used in the model) and an SATM score of 670. Calculate the predicted value of SATM under the model and the residual for this student. (2 points)

$$\hat{\text{satm}} = 426 + 26(9) + 1.64(10) - 7.5(7) = 623.9$$

$$e = 670 - 623.9 = 46.1$$

Question 5 continued

Both of these questions still refer to the regression on the previous page.

(c) A 95% confidence interval for the true HSE coefficient is $(-15.9, 0.9)$. Your colleague explains that this means “If this experiment were repeated many times, in approximately 95% of the repetitions, the true value of the parameter would occur between -15.9 and 0.9 ”. Replace the underlined portion at the end of this incorrect explanation, leaving the rest unchanged, to make it correct. (1 point)

“... within the CI computed for that repetition.”

(d) In the mathematical model for the regression, the error terms ϵ_i all have the same standard deviation σ . From the output, what is the estimate of σ and how many degrees of freedom are associated with it? (2 points)

$s = 76.96$ on $n - p - 1 = 224 - 3 - 1 = 220$ degrees of freedom.

Question 6 (2 points) Here is a simple random sample from a normal distribution with unknown mean and unknown standard deviation:

102 122 101 114 96

(a) Find the sample standard deviation s . (1 point)

$\bar{x} = 107$. Therefore,

$$s^2 = \frac{(-5)^2 + 15^2 + (-6)^2 + 7^2 + (-11)^2}{4} = \frac{456}{4} = 114.$$

This gives $s = \sqrt{114} = 10.677$.

(b) Give a 95% confidence interval for the mean μ , showing all of your work. (1 point)

The confidence interval is $\bar{x} \pm t^*(s/\sqrt{n})$. On 4 degrees of freedom, $t^* = 2.7764$. Thus, the CI is

$$107 \pm 2.7764 \left(\frac{10.677}{\sqrt{5}} \right) = 107.0 \pm 13.3$$

Question 7 (3 points) Imagine that a fair coin will be flipped and a fair 6-sided die will be rolled.

(a) How many events are in the sample space for this simple experiment? (1 point)

This question contains a mistake; “events” should be “outcomes”. There are 12 outcomes. If you’re curious, there are $2^{12} = 4096$ events (in general, if there are n outcomes, there are 2^n events).

Everyone in the class received the point for this question automatically.

(b) Let A be the event that the coin is heads and B the event that the die is 1 or 2. Find $P(A \text{ or } B)$, showing your work. How is the independence of A and B used? (2 points)

Independence tells us that $P(A \text{ and } B) = P(A)P(B)$. Therefore,

$$P(A \text{ or } B) = P(A) + P(B) - P(A)P(B) = \frac{1}{2} + \frac{1}{3} - \frac{1}{6} = \frac{2}{3}$$

Incidentally, many of you found this probability by counting the outcomes in the event A and B . This assumes equally likely outcomes. The outcomes are in fact equally likely, *but only because A and B are independent*. (This is a fairly deep fact that you may want to ponder.) Thus, the independence of the events is fundamentally necessary in order to answer this question.

Question 8 (5 points) Suppose that the heights of 25-year-old women are approximately normally distributed with unknown mean and standard deviation $\sigma = 2.5$ inches. We draw a SRS from the population of 25-year-old women with $n = 10$. We wish to test the null hypothesis $H_0 : \mu = 66$ against the one-sided alternative $H_a : \mu < 66$. Assume that $\alpha = 0.05$.

(a) What is the standard deviation of the sample mean \bar{x} ? (1 point)

$$\frac{\sigma}{\sqrt{n}} = \frac{2.5}{\sqrt{10}} = 0.7906$$

(b) What is the power of the test against the alternative $H : \mu = 65$? Show all of your work. (2 points)

We reject H_0 when $z < -1.645$ (found in Table A). But $z = (\bar{x} - 66)/(.7906)$. Solving for \bar{x} , we reject when $\bar{x} < 64.7$.

By definition, the power against a given alternative is the probability that we reject if we assume that alternative to be the truth. Thus, the power is $P(\bar{x} < 64.7)$ under the assumption that the true mean is 65. Computing this, we get

$$P(\bar{x} < 64.7) = P\left(\frac{\bar{x} - 65}{.7906} < \frac{64.7 - 65}{.7906}\right) = P(Z < -.380) = 0.3522$$

(c) If we observe $\bar{x} = 64$, find the z-statistic and the p-value of the test. (2 points)

$$z = \frac{\bar{x} - 66}{.7906} = \frac{64 - 66}{.7906} = -2.53$$

$$P(Z < -2.53) = 0.0057$$