

MATH 338

EXAM 4

THURSDAY, AUGUST 3, 2017

Your name: _____

Your scores (to be filled in by Dr. Wynne):

Problem 1: ____/23.5

Problem 2: ____/11.5

Problem 3: ____/9

Total: ____/44

You have 60 minutes to complete this exam. This exam is closed book and closed notes with the exception of your formula sheet.

For full credit, show all work except for final numerical calculations (which can be done using a scientific calculator).

1. In a 1936 paper, the great statistician (and geneticist!) Ronald Fisher reexamined Gregor Mendel's famous pea plant experiments. In a hybrid cross, according to Mendel's Law of Independent Assortment, there is a $3/4$ chance of an offspring having the dominant phenotype and a $1/4$ chance of having the recessive phenotype (a phenotype is an observed characteristic determined by genetics).

A) [1.5 pts] If a sample of 7324 seeds is observed from a hybrid cross, what is the exact distribution of the number of seeds with a dominant phenotype (shape/type of distribution, and parameters)?

B) [2.5 pts] Find the mean and standard deviation of the proportion of seeds with a dominant phenotype, when a sample of 7324 seeds is observed.

C) [3 pts] To make sure that a pea plant was a hybrid, Mendel investigated a sample of 10 offspring of a self-fertilization. If any of them had a recessive phenotype, Mendel classified the plant as a hybrid. If the Law of Independent Assortment holds, what is the probability of misclassifying a hybrid as not a hybrid?

D) [6 pts] Out of 7324 seeds from a hybrid cross, Mendel observed 5474 round (dominant) seeds and 1850 wrinkled (recessive) seeds. Construct and interpret a 95% confidence interval for the true proportion of round seeds in the population of pea plants.

E) [2 pts] Based on your work in part (D), do you agree with Fisher's statement that "the deviation from the expected 3:1 [ratio] is less than its standard error of random sampling" (the standard error of \hat{p})? Show an appropriate comparison to justify your answer.

F) [2 pts] If we had no idea about the true population proportion of round seeds, how many seeds would we need to sample for our confidence interval in part (D) to be guaranteed to have a margin of error no greater than 3 percentage points?

G) [6.5 pts] In a dihybrid cross, we expect that the offspring will show a 9:3:3:1 ratio of phenotypes (expected proportions are 9/16, 3/16, 3/16, 1/16). In Mendel's experiments, he observed 315 round, yellow seeds; 108 round, green seeds; 101 wrinkled, yellow seeds; and 32 wrinkled, green seeds. Perform an appropriate hypothesis test to assess whether this generation fits the expected ratio.

2. In 2017, a Polish group sent 360 academic journals the resume of a fake researcher (whose name translates as “Dr. Fraud”), who asked to become an editor. Some of these journals were on a controversial list of “Predatory” journals that will publish anything for a fee, and some weren’t.

A) [2 pts] Fill in the blanks in the two-way table below. The column variable is “What did the journal do to the request?” and the row variable is, “Did the journal appear on the Predatory journals list?”.

| | Accepted | Denied | No Response | Total |
|-------|----------|--------|-------------|-------|
| Yes | 41 | | 69 | 130 |
| No | | | | 230 |
| Total | 48 | 108 | 204 | 360 |

B) [1 pt] If a journal is selected at random from the 360, what is the probability it did not respond?

C) [2 pts] Using the table, estimate the probability that a Predatory journal would accept the request.

D) [6.5 pts] Is there a relationship between a journal’s “Predatory” status and the journal’s decision regarding the fake researcher? Perform an appropriate hypothesis test (the table below may help you organize) and state your conclusion with respect to the real-world question asked.

| | Accepted | Denied | No Response | Total |
|-------|----------|--------|-------------|-------|
| Yes | | | | 130 |
| No | | | | 230 |
| Total | 48 | 108 | 204 | 360 |

3. A certain biomarker test will detect breast cancer in 95% of post-menopausal women who have it, but it will also detect breast cancer in 1% of women who don't have it.

A) [1 pt] What is the sensitivity of this test?

B) [1 pt] What is the specificity of this test?

C) [5 pts] As of 2011, the accepted prevalence (base rate) of breast cancer in post-menopausal women was 0.4% (1 in 2,500 women). Calculate the positive predictive value of the test.

D) [2 pts] A group of researchers evaluated this biomarker test on a sample of post-menopausal women judged to be at high risk for breast cancer, and (correctly) obtained a positive predictive value of over 99%. Why might there be a discrepancy between your answer in part (C) and the researchers' results?

Extra Space. The tables below show a number of critical values z for the standard normal variable $Z \sim N(0, 1)$ and the corresponding cumulative proportions, corresponding to $P(Z \leq z)$.

| z-score | Cumulative Proportion |
|---------|-----------------------|
| -3.00 | 0.0013 |
| -2.50 | 0.0062 |
| -2.00 | 0.0228 |
| -1.65 | 0.0495 |
| -1.28 | 0.1003 |
| -1.00 | 0.1587 |
| -0.67 | 0.2514 |

| z-score | Cumulative Proportion |
|---------|-----------------------|
| 0.67 | 0.7486 |
| 1.00 | 0.8413 |
| 1.28 | 0.8997 |
| 1.65 | 0.9505 |
| 2.00 | 0.9772 |
| 2.50 | 0.9938 |
| 3.00 | 0.9987 |

Refer to the following table for z^* critical values for confidence intervals:

| | C = 0.90 (90%) | C = 0.95 (95%) | C = 0.98 (98%) | C = 0.99 (99%) | C = 0.999 (99.9%) |
|--------------|----------------|----------------|----------------|----------------|-------------------|
| z^* values | 1.645 | 1.960 | 2.326 | 2.576 | 3.291 |

For a two-sided hypothesis test, use the column corresponding to $C = 1 - \alpha$

For a one-sided hypothesis test, use the column corresponding to $C = 1 - 2\alpha$

Refer to the following table for χ^2 critical values:

| Degrees of freedom | $\alpha = 0.05$ | $\alpha = 0.01$ |
|--------------------|-----------------|-----------------|
| 1 | 3.84 | 6.63 |
| 2 | 5.99 | 9.21 |
| 3 | 7.81 | 11.34 |
| 4 | 9.49 | 13.28 |
| 5 | 11.07 | 15.09 |
| 6 | 12.59 | 16.81 |

The rest of this space to be used for extra work: