STAT 3332 Statistics for Life Sciences

Fall 2006 — Final Examination — 11/28/06

• Do **not** turn in the question sheets. **Only the answer sheet** will be accepted.

• There are **32** questions, with **3** points per question plus **4** points for neatness of name and signature on the answer sheet. Total points: **100**.

• Allotted time: **120** minutes.

• For each question, choose the **best answer**. Explanations for any question (clearly indicate which) may be provided on the back of the answer sheet and may count for **partial credit** in the case of a wrong answer.

---

1. The sample space of outcomes for a roll of a die is $S = \{1, 2, 3, 4, 5, 6\}$. For the two events $A = \{1, 2, 3\}$ and $B = \{3, 4, 5, 6\}$ in S,

* (A) $A \cup B = S$ (B) $A \cap B = \emptyset$ (C) $\overline{A} = B$ (D) $\overline{B} = S$

---

2. Let events A, B, and C have respective probabilities 1/2, 1/3, and 1/4, and let $A \cap B$, $B \cap C$, and $A \cap C$ have respective probabilities 1/6, 1/10, and 1/16. Which two of A, B, and C are *independent* events?

* (A) A and B (B) B and C (C) A and C

---

3. (Continuation) For A and B as above, $P(A \cup B) =$

(A) $\frac{1}{2} + \frac{1}{3}$ (B) $\frac{1}{2} \times \frac{1}{3}$ (C) $\frac{1/2}{1/3}$ * (D) $\frac{1}{2} + \frac{1}{3} - (\frac{1}{2} \times \frac{1}{3})$

---

4. Let random variable X have a continuous probability distribution with density function $f(x)$. Then $P(-10 \leq X \leq 20) =$

(A) $f(20) - f(-10)$ * (B) $\int_{-10}^{20} f(x)dx$ (C) $\int_{-\infty}^{\infty} f(x)dx$ (D) $f(20)$

---

5. (Continuation) And $P(X = 1) =$

(A) $f(1)$ (B) 1 * (C) 0 (C) -1

---

6. For sampling of size 30 from $N(\mu, \sigma^2)$ with both parameters unknown, a 90% confidence interval for μ based on the sample mean $\overline{X}$ and the sample standard deviation s is

(A) $\overline{X} \pm s$ (B) $\overline{X} \pm 1.96\,s$ (C) $\overline{X} \pm t_{29,0.30} \frac{s^2}{29}$ * (D) $\overline{X} \pm t_{29,0.95} \frac{s}{\sqrt{30}}$

---

7. For sampling of size 30 from any population with mean μ and variance σ^2, the sample variance $s^2 = (29)^{-1} \sum_{i=1}^{30} (X_i - \overline{X})^2$ has mean $E(s^2) =$

(A) μ * (B) σ^2 (C) $\frac{\sigma^2}{\sqrt{30}}$ (D) 31

---

8. Among 40 randomly selected households in a neighborhood, 15 ($\hat{p} = 0.375$) favor letting Wal-Mart put a store in the neighborhood. To form a confidence interval for the actual proportion p of households favoring this action, approximate the exact distribution of $\hat{p}$ by

(A) $N(0, 1)$ (B) $N(0.375, 1)$ (C) $N(0.375, 0.375)$ * (D) $N\left(0.375, \frac{0.375 \times 0.625}{40}\right)$

---

9. Random numbers are used in statistics for

(A) sample selection (B) avoiding bias * (C) both of these

---

10. In hypothesis testing at level of significance $\alpha = .01$ using a p-value, the hypothesis H_0 would be rejected if the p-value for the data satisfied

* (A) p-value = 0.00001 (B) p-value > 0.01 (C) p-value > 0.05

---

11. In an opthalmology study of an eye drop, each participant randomly received either the active drug (A) or a placebo (P) in the left (L) eye and the other of these in the right (R) eye. Effectiveness was scored for each eye of each participant. To best analyze the data by taking differences, compute for each participant

(A) $L - R$ (B) $L - A$ (C) $L - P$ (D) $P - R$ * (E) $P - A$

---

12. *Cardiovascular Disease.* Suppose for a population of persons it is known that

P(person is hypertensive) = .02,
P(person is normal) = .98,
P(chest pain | hypertensive person) = .10,
P(chest pain | normal person) = .05.

Then the probability that a randomly selected person has chest pain is

(A) .10 + .05 = .15 (B) .98 × .05 = .049 * (C) .98 × .05 + .02 × .10 = .051

---

13. (Continuation) The conditional probability P(normal | chest pain) is

* (A) $\frac{.049}{.051} = 0.96$ (B) .002 (C) .049

---

14. For the sample space $S = \{\text{HH, HT, TH, TT}\}$ and probability function $P(s)$ on outcomes s in S, define the random variable $X(s)$ = number of Heads in s. The induced probability $P(X = 1)$ is given by

(A) $P(\text{HH}) + P(\text{HT})$ * (B) $P(\text{HT}) + P(\text{HT})$ (C) $P(\text{HT}) + P(\text{TT})$

---

15. For the discrete probability distribution

x	0	1	2	3
$p(x)$	0.2	0.3	0.3	0.2

the *expected value* $E(X)$ is

(A) 0 * (B) 1.5 (C) 3

---

16. Georgia, Audrey, Sarah, and Cassandra would like to go for a ride with Bob in his new car. He plans to take 2 passengers plus his dog. In how many ways can Bob select 2 passengers from these 4 persons?

(A) 2 (B) $\binom{4}{3} = 4$ * (C) $\frac{4!}{(2!)^2} = 6$ (D) $4 \times 2 = 8$

---

17. (Continuation) The dog sits in the front. In how many ways can the 2 selected passengers be assigned to the seats "left rear" and "right rear"?

(A) 0 * (B) $\binom{2}{1}$ (C) 1

---

18. The probability distribution for the number of successes in 3 independent success-failure trials with success probability 1/3 is the binomial$(3, 1/3)$ distribution:

x	0	1	2	3
$p(x)$	$\frac{8}{27}$	$\frac{12}{27}$	$\frac{6}{27}$	$\frac{1}{27}$

Given that *at least one success occurs* in the 3 trials, what is the conditional probability that the number of successes is exactly 2?

(A) 0 (B) $\frac{5}{9}$ * (C) $\frac{6}{19}$ (D) 1 (E) ∞

---

19. Suppose that a rocket ship has 2,000,000 components performing independently, each with a very small chance $\frac{1}{1{,}000{,}000}$ of failing during a mission. Then the number of failures occurring in a mission is *binomial* with $n = 2{,}000{,}000$ and $p = \frac{1}{1{,}000{,}000}$. Approximation of this distribution by *Poisson* with mean 2 is justified because

(A) $np = 2$ (B) np^2 is small * (C) both of these

---

20. For sampling from any population with mean μ and variance σ^2,

(A) the sample mean $\overline{X}$ has mean μ

(B) the sample mean $\overline{X}$ has variance $\frac{\sigma^2}{n}$

(C) the standardized sample mean $\frac{\overline{X}-\mu}{\sigma/\sqrt{n}}$ has mean 0 and variance 1

* (D) each of (A), (B), (C) is true

(E) none of (A), (B), (C) is true

---

21. Consider testing $H_0 : \mu = 33$ (my age) versus $H_1 : \mu > 33$ with σ unknown, based on the sample mean $\overline{X}$. To compute the p-value we need to know the distribution of $\frac{\overline{X}-33}{s/\sqrt{n}}$

* (A) when H_0 is true (B) when H_1 is true (C) in each of these cases

---

22. An interval estimate of $\mu_1 - \mu_2$ is to be constructed based on $\overline{X}_1 - \overline{X}_2$ and sample variances s_1^2 and s_2^2 from independent samples of equal size from $N(\mu_1, \sigma^2)$ and $N(\mu_2, \sigma^2)$, with σ^2 unknown. The most appropriate estimator of σ^2 to use is

(A) s_1^2 (B) s_2^2 (C) $\frac{s_1^2}{1} + \frac{s_2^2}{2}$ * (D) $\frac{1}{2}(s_1^2 + s_2^2)$

---

23. Consider the 3×2 contingency table (with row, column, and grand totals in **bold**)

19	14	9	**42**
5	10	6	**21**
24	**24**	**15**	**63**

For the $(1, 1)$ cell with *observed count* $O_{11} = 19$, the *expected count* is $E_{11} =$

(A) 19 (B) $\frac{24+42}{63}$ (C) $\frac{24-42}{63}$ (D) 63 * (E) $\frac{24\times42}{63}$

---

24. (Continuation) For testing homogeneity or independence, the test statistic $\sum_{cells} \frac{(O_{ij}-E_{ij})^2}{E_{ij}}$ has approximate distribution *chi square* with degrees of freedom

(A) $6 - 1 = 5$ * (B) $(3 - 1) \times (2 - 1) = 2$ (C) 0 (D) 6 (E) $63 - 1 = 62$

25. The linear regression model for the relation between X and Y is described by

(A) $Y = \alpha + \beta X + \varepsilon$ with $E(\varepsilon) = 0$.

(B) $E(Y \mid X = x) = \alpha + \beta x$

* (C) both of these

(D) neither of these

---

26. From a set of hematologic data for patients with aplastic anemia, a regression line relating number of lymphocytes (per mm^2) (Y) to % reticulocytes (X) was fitted:

$$Y = 1894.8 + 112.1X.$$

For two subjects whose reticulocytes differ by 5%, what is the expected difference in their numbers of lymphocytes?

(A) 112.1   * (B) 5×112.1 (C) $1894.8 + 5 \times 112.1$ (D) $5 \times (1894.8 + 112.1)$

---

27. For the following data set of 9 values listed in increasing order, give the *median.*

-120 -12 4   23   * 500   600   999   12,021   56,001

---

28. (Continuation) *Adding* the constant $c = 10$ to each value in this data set adds 10 to the

* (A) mean   (B) interquartile range   (C) standard deviation   (D) each of these

---

29. A *boxplot* uses the

(A) median   (B) 25th percentile   (C) interquartile range   * (D) each of these

---

30. *Multiplying* each value in a data set by the constant $c = 20$ changes

(A) $\overline{X}$ (B) s^2 * (C) each of these   (D) neither of these

---

31. The correlation $\rho_{XY} = \text{Cov}(X, Y)/\sigma_X \sigma_Y = E[(X - \mu_X)(Y - \mu_Y)]/\sigma_X \sigma_Y$ measures what feature of the relationship between X and Y? (*write answer – linearity – on answer sheet*)

---

32. Consider two independent samples of sizes n_1 and n_2, respectively, from populations with possibly differing means μ_1 and μ_2 but equal variance $\sigma^2 = 25$. Using the respective sample variances s_1^2 and s_2^2, the best estimator of σ^2 is $\hat{\sigma}^2 =$

(A) $\frac{s_1^2 + s_2^2}{2}$ (B) $\max\{s_1^2, s_2^2\}$ * (C) $\frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2}$ (D) $\frac{1}{s_1} + \frac{1}{s_2}$

- When you are finished, *please hand in* **only** *the answer sheet.*
- *KEEP THE QUESTION SHEETS.*
- Please **depart quietly**, and *leave the vicinity before discussing the test.*

Thank you for your attention to these details.