STAT 3332 Statistics for Life Sciences

Spring 2008 — Final Examination — May 5, 2008

• Do **not** turn in the question sheets. **Only the answer sheet** will be accepted.
• There are **32** questions. Total points **100**: **3** per question, and **4** for *printing* name on answer sheet.
• Allotted time: **120** minutes.
• For each question, choose or give the **best answer**. Explanations (clearly indicate for which question) provided **neatly** on *the margin or back of the answer sheet* may count for **partial credit** if needed.

---

1. The sample space of outcomes for a family of three children is

$S = \{\text{bbb, bbg, bgb, bgg, gbb, gbg, ggb, ggg}\}$,

where b denotes "boy" and g denotes "girl". For the events $A = \{\text{bbb, bbg, ggg}\}$ and $B = \{\text{bbg, bgb, gbb}\}$,

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

---

2. (Continuation) Let probability function P attach equal probability 1/8 to the 8 outcomes in S. The conditional probability of A given B is

* (A) $P(A \mid B) = \frac{P(A \cap B)}{P(B)} = \frac{1/8}{3/8} = 1/3$

(B) $P(A \mid B) = \frac{P(A \cup B)}{P(A)} = \frac{5/8}{3/8} = 5/3$

(C) $P(A \mid B) = P(A) = 3/8$

(D) $P(A \mid B) = \frac{P(A)}{P(B)} = \frac{3/8}{3/8} = 1.$

---

3. (Continuation) For the random variable $X(s)$ = "number of girls g in outcome s", the induced probability $P(X = 2)$ is

* (A) $P(\{\text{bgg, gbg, ggb}\})$ (B) $P(\{\text{gg}\})$ (C) $P(\{\text{bgg, gbg, ggb, ggg}\})$ (D) $P(\text{S})$

---

4. Let random variable X have a continuous probability distribution with density function $f(x)$. Then $f(x)$ represents

(A) $P(X = x)$ (B) $P(X > x)$ (C) $P(X < x)$ * (D) none of these

---

5. The *median* of the 9 values

$-12 \quad 4 \quad 23 \quad 50 \quad 600 \quad 2,999 \quad 12,021 \quad 56,001 \quad 100,000$

is

(A) -12 (B) 100,000 (C) 4 (D) 56,001 * (E) 600

---

6. (Continuation) *Adding* the constant $c = 7$ to each value in this data set adds 7 to the

(A) IQR (B) standard deviation * (C) 3rd quartile (D) all of these

---

7. A scale-free measure of *skewness* is

(A) 2× (standard deviation) (B) 1.5× IQR (C) median (D) IQR * (E) $\frac{\text{mean} - \text{median}}{\text{standard deviation}}$

---

8. Suppose that an animal's genome has $n = 30,000$ genes, each independently having a very small chance $p = \frac{1}{10,000}$ of experiencing a mutation today. Then the number of mutations occurring today has approximate probability distribution distribution

* (A) Poisson(np) (B) Exponential(n/p) (C) Normal(n, p) (D) Geometric(p)

---

9. 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 1/5, 1/12, and 1/8. Which two events are *not* independent?

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

---

10. If you have just bought a dozen (12) doughnuts of all different kinds, in how many ways can you select 3 for eating now?

(A) $\frac{12}{3}$ * (B) $\frac{12!}{3!9!}$ (C) $\frac{12!}{3!}$ (D) 12×9

---

11. Suppose that independent random samples from two distributions F and G are

From F:	8	20	21	27	42
From G:	19	26	28	60	61

For testing $H_0 : F = G$, the *Wilcoxon rank sum test statistic* W is

(A) $8 + 20 + 21 + 27 + 42$

* (B) $1 + 3 + 4 + 6 + 8$

(C) $1 \times 8 + 3 \times 20 + 4 \times 21 + 6 \times 27 + 8 \times 42$

---

12. 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 (A) and (P) in the right (R) eye. Effectiveness was scored for each eye of each participant. To test the hypothesis of no difference between treatment and placebo, one may apply a one-sample test in this problem by computing for each participant the difference

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

---

13. *Voting.* Suppose for a population of voters it is known that

P(voter votes for Clinton) = .7
P(voter is female) = .6
P(voter votes for Clinton | voter is female) = .8
P(voter votes for Clinton | voter is male) = .4

Then the probability that a randomly selected voter is both *male* and *votes for Clinton* is

(A) $.8 + .4 = 1.2$ (B) $.6 \times .8 = .48$ * (C) $.4 \times .4 = .16$

---

14. (Continuation) The conditional probability P(voter is male | voter votes for Clinton) is

* (A) $\frac{.16}{.16+.48}$ (B) $.4 \times .4 = .16$ (C) $.7 \times .4 = .28$

---

15. For the discrete probability distribution

x	0	1	2	3
$p(x)$	0.1	0.2	0.3	0.4

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

(A) $\frac{1+2+3}{3}$

(B) $\frac{1\times0.2+2\times0.3+3\times0.4}{4}$

* (C) $1 \times 0.2 + 2 \times 0.3 + 3 \times 0.4$

---

16. Let random variable X have the **Standard Normal** distribution, $N(0, 1)$. Then $P(0 \le X \le 1)$ has *approximate value*

(A) $\frac{1}{2}$ * (B) $\frac{1}{3}$ (C) 0 (D) -0.025 none of these

---

17. The probability distribution for the number of successes in 3 independent success-failure trials with success probability 1/4 is the binomial(3, 1/4) distribution:

x	0	1	2	3
$p(x)$	$\frac{27}{64}$	$\frac{27}{64}$	$\frac{9}{64}$	$\frac{1}{64}$

Given that *at least 1 success occurs* in the 3 trials, what is the conditional probability that the number of successes is *exactly 2*?

(A) 0 (B) $\frac{9}{64}$ * (C) $\frac{9/64}{37/64} = \frac{9}{37}$ (D) $\frac{2+3}{9+1} = \frac{1}{2}$ (E) 1

---

18. In testing a null hypothesis H_0, suppose that a p-value of 0.004 is obtained. Then the hypothesis H_0 would be rejected at level of significance α for

* (A) $\alpha = 0.01$ (B) $\alpha = 0.001$ (C) $\alpha = 0.0001$ (D) $\alpha = 0.00001$

---

19. Consider testing $H_0 : \mu = 1$ versus $H_1 : \mu > 2$ for a Normal population with σ unknown, based on the sample mean $\overline{X}$ and the sample standard deviation s for a sample of size $n = 100$. To compute the p-value we need to use the H_0-distribution of $\frac{\overline{X}-1}{s/10}$. The needed distribution is

* (A) t with 99 degrees of freedom (B) $N(0,1)$ (C) chi-square with 100 degrees of freedom

---

20. A hypothesis about $\mu_1 - \mu_2$ is to be tested based on $\overline{X}_1 - \overline{X}_2$ and sample variances s_1^2 and s_2^2 from independent samples of equal size $n_1 = n_2 = 15$ from $N(\mu_1, \sigma_1^2)$ and $N(\mu_2, \sigma_2^2)$, with σ_1^2 and σ_2^2 unknown. A p-value would be based on the t distribution with degrees of freedom given by

(A) 28 * (B) Satterthwaite's approximation (C) 15

---

21. For sampling of size 100 from $N(\mu, 25)$, with sample mean $\overline{X}$ and sample standard deviation s, a 95% confidence interval for μ is given by

(A) $\overline{X} \pm 1.96\, s$ (B) $\overline{X} \pm \frac{s}{10}$ (C) $\overline{X} \pm \frac{s^2}{99}$ * (D) $\overline{X} \pm 1.96\, \frac{5}{10}$

---

22. 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) $\mu + 1.96\,\sigma$ (C) $\frac{\sigma^2}{\sqrt{30}}$ * (D) σ^2

---

23. Consider two independent samples of equal size n 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

* (A) $\frac{s_1^2 + s_2^2}{2}$ (B) $\max\{s_1^2, s_2^2\}$ (C) s_1^2 (D) $\min\{s_1^2, s_2^2\}$ (E) $\sqrt{s_1^2 \times s_2^2}$

---

24. Consider *categorical data* in the 3×2 contingency table (with row totals, column totals, and the grand total shown in **bold**)

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

For the (2, 2) cell with *observed count* $O_{22} = 10$, the *estimated expected count* is $\widehat{E}_{22} =$

10 $\frac{21+24}{63}$ $\frac{24-21}{63}$ * $\frac{21 \times 24}{63}$ 63

---

25. (Continuation) For testing homogeneity or independence, the test statistic $\sum_{cells} (O_{ij} - \widehat{E}_{ij})^2 / \widehat{E}_{ij}$ has approximate distribution *chi square* with degrees of freedom (encircle)

$6 - 1 = 5$ * $(3-1) \times (2-1) = 2$ 0 6 63

---

26. (Continuation) If the column totals **24**, **24**, and **15** are *random*, the statistic $\sum_{cells} \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$ may be used for testing

* (A) homogeneity of 3 populations (B) independence of 2 variables (C) normality

---

27. The Siegel-Tukey test statistic

(A) tests for difference in the spreads of two populations

(B) has the same H_0-distribution as the Wilcoxon rank sum statistic for the same data

* (C) satisfies both (A) and (B)

28. 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

---

29. The mite *Demodex folliculorum* is less than 1 mm in length and lives in the hair follicles of humans, including those of the eyelashes. A biomedical scientist hypothesized that the number of these would increase with a person's age. For 10 subjects of different ages, 25 eyelashes were plucked at random from each eye and the number of mites counted. The data:

mite count	5	13	16	14	18	23	20	32	29	28
age	3	6	9	12	15	18	21	24	27	30

The regression line *mite count* = 5.7 + 0.9 *age* was fitted. what is the approximate average increase in the number of mites found in 50 eyelashes, for each additional year of age?

(A) About 6 * (B) About 1 (C) About 6 + 1 = 7 (D) About 6 − 1 = 5

---

30. The correlation between X and Y, $\rho_{XY} = \mathrm{Cov}(X,Y)/\sigma_X\sigma_Y = E[(X-\mu_X)(Y-\mu_Y)]/\sigma_X\sigma_Y$,

(A) always lies between 0 and 1

(B) measures nonlinearity in the relationship between X and Y

* (C) measures dependence in the relationship between X and Y

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

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

---

31. In fitting a linear regression model $E(Y \mid X = x) = \alpha + \beta x$ by least squares with data (X_i, Y_i), the estimates of α and β are chosen to minimize

(A) the sum of squared fitted values $\{\widehat{Y}_i, 1 \le i \le n\}$

* (B) the sum of squared residuals $\{Y_i - \widehat{Y}_i, 1 \le i \le n\}$

(C) the sum of squared deviations $\{Y_i - \overline{Y}, 1 \le i \le n\}$

---

32. For measuring the goodness of fit of a linear regression model, one may use

(A) The square of the sample correlation

(B) The ratio of the sum of squares for regression to the total sum of squares

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

(D) Neither (A) nor (B)

- 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.