STAT 3332 Statistics for Life Sciences

Spring 2009 — Final Examination — May 11, 2009

- 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, with b = "boy" and g = "girl", is

S = {bbb, bbg, bgb, bgg, gbb, gbg, ggb, ggg}.

For events A = {bbb, bbg, bgb,bgg} and B = {gbb,gbg, ggb, ggg}, which of the following is *not* true?

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

---

2. (Continuation) Let probability function P attach probability 1/8 to each outcome in S. Let C = {ggg}. The conditional probability of C given B is

(A) $P(C \mid B) = \frac{P(B \cup C)}{P(C)} = \frac{4/8}{1/8} = 4$

* (B) $P(C \mid B) = \frac{P(B \cap C)}{P(B)} = \frac{1/8}{4/8} = 1/4$

(C) $P(C \mid B) = P(C) = 1/8$

(D) $P(C \mid B) = P(B) \times P(C) = 4/8 \times 1/8 = 1/16$

---

3. (Continuation) Let A = {bbb, bbg}, B = {bbb, bgg}, and C = {bbb, ggg, ggb, gbg}. Which two events are *not* independent?

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

---

4. (Continuation) Define the random variable $X(s)$ = "number of girls g in outcome s". Then $P(X \geq 2)$ is

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

---

5. Let random variable X have a *continuous* probability distribution with *density* function $f(x)$. Then $P(X \geq 2)$ is

(A) $f(2)$ (B) $1 - f(2)$ * (C) $\int_2^\infty f(x)\,dx$ (D) none of these

---

6. The *median* of the 8 values

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

is

* (A) 325 (B) 50 (C) 600 (D) none of these

---

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

(A) IQR (B) standard deviation (C) MAD * (D) none of these

---

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

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

---

9. 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 mean

(A) 0 * (B) 3 (C) 10,000 (D) 30,000

---

10. If you have just bought a banana, a sandwich, and a cookie, in how many ways can you choose an order in which to eat them?

(A) $\binom{3}{3} = 1$ * (B) $3! = 6$ (C) 3 (D) $3 \times 3 \times 3 = 27$

---

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

From F: 8 20 21 27 42

From G: 9 19 26 28 60 61

For testing H_0 : "the medians of F and G are equal", the *Wilcoxon rank sum test statistic* W is

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

* (B) $1 + 4 + 5 + 7 + 9$

(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 H_0 : "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. *Polygraph (Lie-Detector) Instrument*. Suppose it is known that

P(subject is lying) = 0.2
P(polygraph says "lying" | subject is lying) = 0.8
P(polygraph says "lying" | subject is truthful) = 0.2

Then the probability that the polygraph says "lying" is

(A) $.2 \times .8 = .16$ * (B) $.2 \times .8 + .8 \times .2 = .32$ (C) $.2 \times .2 + .8 \times .8 = .68$

---

14. (Continuation) The probability that the polygraph is correct is

(A) $.2 \times .8 = .16$ (B) $.2 \times .8 + .8 \times .2 = .32$ * (C) $.2 \times .8 + .8 \times .8 = .8$

---

15. (Continuation) The conditional probability P(subject is lying | polygraph says "lying") is

(A) 1 (B) 0 * (C) 0.5

---

16. For the discrete probability distribution

x	1	2	3
$p(x)$	0.3	0.3	0.4

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

(A) $\frac{1+2+3}{3} = 2.0$

(B) $\frac{1\times0.3+2\times0.3+3\times0.4}{3} = 0.7$

* (C) $1 \times 0.3 + 2 \times 0.3 + 3 \times 0.4 = 2.1$

---

17. Let random variable X have the **Standard Normal** distribution, $N(0, 1)$. Then $P(0 \le X \le 3)$ has *approximate value*

* (A) $\frac{1}{2}$ (B) $\frac{1}{9}$ (C) 0 (D) -0.025

---

18. 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 2 successes occur* in the 3 trials, what is the conditional probability that the number of successes is *exactly 2*?

(A) 0 (B) $\frac{9}{64}$ * (C) $\frac{9/64}{10/64} = \frac{9}{10}$ (D) $\frac{2+3}{9+1} = \frac{1}{2}$ (E) 1

---

19. In testing a null hypothesis H_0, suppose that a p-value of 0.04 is obtained. Then the hypothesis H_0 would be rejected at level of significance α for

* (A) $\alpha = 0.05$ (B) $\alpha = 0.01$ (C) $\alpha = 0.001$ (D) $\alpha = 0.0001$

---

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 esimated by s_1^2 and s_2^2. A p-value would be based on the t distribution with degrees of freedom given by

(A) 28 * (B) Satterthwaite's approximation using s_1^2 and s_2^2 (C) 15

---

21. For sampling of size 100 from $N(\mu, \sigma^2)$, 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) $\overline{X} \pm 1.96\, \frac{s}{10}$

---

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 $(2, 3)$ cell with *observed count* $O_{23} = 6$, the *estimated expected count* is $\widehat{E}_{23} =$

(A) 6 (B) $\frac{21+15}{63}$ (C) $\frac{21-15}{63}$ * (D) $\frac{21 \times 15}{63}$ (E) 63

---

25. (Continuation) For testing homogeneity or independence, the test statistic $\sum_{\text{cells}} (O_{ij} - \widehat{E}_{ij})^2 / \widehat{E}_{ij}$ 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

---

26. (Continuation) If the column totals and row totals are *random* with the total sample size (**63**) fixed, the statistic $\sum_{\text{cells}} \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$ may be used for testing

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

---

27. The Kruskal-Wallis test statistic

(A) tests for differences in the variances of several populations

* (B) tests for differences in the medians of several populations

(C) assumes the data are normally distributed

---

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, the same number N of 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. If your age is 22, what is your best guess of the mite count for N eyelashes plucked at random from each of your eyes?

(A) About 5 (B) About 10 (C) About 15 (D) About 20 * (E) About 25

---

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

(A) always lies between 0 and 1

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

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

(D) equals 0 only if X and Y are independent

(E) none of (A), (B), (C), and (D) 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 mean of the residuals

(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.