

Giảng viên tổng hợp đề: (Chữ ký và Họ tên)	Ngày ra đề 10/08/2022	Người phê duyệt: (Chữ ký, Chức vụ và Họ tên)	Ngày duyệt đề:
--	-----------------------	--	----------------

(phần phía trên cần che đi khi in sao đề thi)

 TRƯỜNG ĐH BÁCH KHOA – ĐHQG-HCM KHOA KHOA HỌC ỨNG DỤNG	ĐỀ THI CUỐI KỲ		Học kỳ/năm học	3	2021-2022	
			Ngày thi	21/08/2022		
	Môn học	XÁC SUẤT THỐNG KÊ				
	Mã môn học	MT2013				
	Thời lượng	100 phút	Mã đề	DT		
Ghi chú: <ul style="list-style-type: none">- Đề thi gồm 2 trang A4.- Sinh viên được sử dụng máy tính bỏ túi; các bảng tra số; tài liệu giấy được in hoặc photo ở khổ A4, có ghi họ tên sinh viên (không sử dụng tài liệu được viết tay).- Sinh viên không được trao đổi tài liệu trong phòng thi.- Sinh viên không làm tròn kết quả trung gian. Các đáp án gần đúng lấy tròn 4 chữ số phần thập phân.- Nộp lại đề thi cùng với bài làm						

Câu hỏi 1 (L.O.2.1): (2 điểm)

Một nhà máy sản xuất linh kiện đã thống kê về số lỗi trên mỗi sản phẩm như sau:

Số lỗi trên 1 linh kiện	0	1	2
Tỉ lệ tương ứng	0.85	0.1	0.05

- a) Nhà máy mới đưa vào sử dụng một thiết bị tự động để kiểm tra số lỗi trên mỗi linh kiện ngay sau khi chúng được sản xuất ra. Bảng dưới đây cho biết hiệu quả hoạt động của thiết bị kiểm tra này:

Số lỗi thực tế trên một linh kiện	Kết luận của thiết bị đối với linh kiện
0	<ul style="list-style-type: none"> • 95% trường hợp kết luận có 0 lỗi • 5% trường hợp kết luận có 1 lỗi
1	<ul style="list-style-type: none"> • 5% trường hợp kết luận có 0 lỗi • 90% trường hợp kết luận có 1 lỗi • 5% trường hợp kết luận có 2 lỗi
2	<ul style="list-style-type: none"> • 20% trường hợp kết luận có 1 lỗi • 80% trường hợp kết luận có 2 lỗi

Nếu một linh kiện được thiết bị kiểm tra kết luận không có lỗi nào thì xác suất linh kiện đó thực sự không có lỗi là bao nhiêu?

- b) Lấy ngẫu nhiên 120 linh kiện từ nhà máy. Gọi Z là biến ngẫu nhiên chỉ tổng số lỗi trên các linh kiện được lấy ra. Tìm kỳ vọng $E(Z)$ và $P(Z \leq 20)$.

Câu hỏi 2 (L.O.2.1): (4 điểm)

Một kỳ nghỉ ngắn ngày cũng có thể làm trẻ em quên mất một phần kiến thức được học trong trường ngay trước khi nghỉ. Ở một trường tiểu học, các thầy cô đã cho mỗi học sinh làm một bài kiểm tra ngay trước kỳ nghỉ và một bài kiểm tra ngay khi học sinh kết thúc kỳ nghỉ và trở lại trường học. Hai bài kiểm có mức độ đánh giá kiến thức tương đương nhau.

Điểm kiểm tra của học sinh là những biến ngẫu nhiên tuân theo phân phối chuẩn.

Dưới đây là điểm số ghi nhận được ở một mẫu gồm 8 học sinh:

Điểm trước kỳ nghỉ.	7	8	9	6	7	6	8	9
Điểm tương ứng sau kỳ nghỉ.	6	8	7.5	5	7.5	6	6.5	8.5

- Với mức ý nghĩa 5%, có thể cho rằng điểm số trung bình của học sinh đã giảm sau kỳ nghỉ hay không?
- Tìm khoảng tin cậy 99% cho điểm trung bình của học sinh trước kỳ nghỉ.
- Biết rằng trong số 92 học sinh được hỏi thì có 40 em trả lời đã tự ôn bài trong kỳ nghỉ. Có thể xem như một nửa số học sinh đã ôn bài trước khi trở lại trường học hay không? Hãy kết luận với mức ý nghĩa 5%.

Câu hỏi 3 (L.O.2.1): (2 điểm)

Một nhà dinh dưỡng học đã đề nghị những người thường xuyên đạp xe ở cùng độ tuổi tham gia vào một thử nghiệm. Họ được chia ngẫu nhiên thành 3 nhóm. Nhóm thứ nhất được yêu cầu ăn uống bổ sung nhiều loại vitamin; nhóm thứ hai có chế độ ăn bổ sung nhiều ngũ cốc giàu chất xơ; nhóm thứ ba được bổ sung nhiều loại khoáng chất. Sau 4 tuần lễ, những người tham gia thực nghiệm sẽ đạp xe liên tục trong 6 giờ. Vận tốc đạp xe trung bình (đơn vị: km/giờ) của mỗi người được ghi nhận như sau:

Nhóm 1	15.6	16.4	17.2	17.1	16.6
Nhóm 2	17.3	16.6	15.5	16.8	17.2
Nhóm 3	16.4	16.9	15.8	16.2	17

Hãy sử dụng phương pháp Anova để so sánh hiệu quả của 3 chế độ dinh dưỡng bổ sung lên thành tích đạp xe của những người tham gia thử nghiệm và kết luận với mức ý nghĩa 5%. Giả thiết rằng các số liệu phù hợp với phương pháp Anova.

Câu hỏi 4 (L.O.2.1): (2 điểm)

Khi đo một số cây thông ba lá trong rừng thông ở Lâm đồng, người ta ghi nhận lại số liệu trong bảng dưới đây. Gọi $X(\text{cm})$ là đường kính thân cây và $Y(\text{m})$ là chiều cao tương ứng.

$X(\text{cm})$	6.2	8.3	10.4	12	13.6	15.5	16
$Y(\text{m})$	3.4	7.5	8.5	9	10.5	11	11

- Tìm hệ số tương quan của mẫu hai chiều và nêu nhận xét.
- Hãy ước tính chiều cao của cây thông có đường kính 10.5 cm.
- Hãy tìm khoảng tin cậy 95% cho hệ số góc của đường hồi quy tuyến tính Y theo X .

--- HẾT ---

ĐÁP ÁN

Câu 1: (1đ+ 1đ)

- a) Gọi T là biến cố linh kiện không có lỗi nào,
và KT là biến cố linh kiện được thiết bị kết luận không có lỗi.
Xác suất cần tìm:

$$P(T / KT) = \frac{P(T * KT)}{P(KT)}.$$

Sử dụng công thức đầy đủ:

$$P(KT) = P(\text{linh kiện 0 lỗi}) \times P(KT / \text{linh kiện 0 lỗi}) + P(\text{linh kiện 1 lỗi}) \times P(KT / \text{linh kiện 1 lỗi}) + P(\text{linh kiện 2 lỗi}) \times P(KT / \text{linh kiện 2 lỗi}).$$

$$= 0.85 \times 0.95 + 0.1 \times 0.05 + 0.05 \times 0 = 0.8125$$

$$P(T * KT) = 0.85 \times 0.95 = 0.8075$$

Đáp số: 0.9938

- b) Gọi X là số lỗi trên một linh kiện bất kỳ. X có bảng phân phối xác suất như đề bài cho.
Gọi X_i là số lỗi trên linh kiện thứ i; $i = 1; 2; \dots; 120$.

Các X_i có cùng phân phối xác suất với X. Xem như các X_i là độc lập nhau.

Như vậy $Z = X_1 + X_2 + \dots + X_{120}$

$$E(Z) = 120 \times E(X) = 120 \times 0.2 = 24$$

$$V(Z) = 120 \times V(X) = 120 \times 0.26 = 31.2$$

Theo định lý giới hạn trung tâm, Z xấp xỉ phân phối chuẩn $N(24; 31.2)$

Xác suất để tổng số lỗi không quá 20:

$$P(Z \leq 20) = \Phi\left(\frac{20.5 - 24}{\sqrt{31.2}}\right) = 0.26556$$

Câu 2: (1.5 đ+ 1đ +1.5đ)

- a) Gọi $\mu_1; \mu_2$ lần lượt là điểm trung bình của học sinh trước và sau kỳ nghỉ.

Gt $H_0: \mu_1 = \mu_2$

Gt $H_1: \mu_1 > \mu_2$

Miền bác bỏ $RR = (t_{0.05}(7); +\infty) = (1.895; +\infty)$

$X_D = X_1 - X_2$

X1	7	8	9	6	7	6	8	9
X2	6	8	7.5	5	7.5	6	6.5	8.5
X _D	1	0	1.5	1	-0.5	0	1.5	0.5

Các đặc trưng mẫu X_D : $n = 8$ $\bar{x}_D = 0.625$ $s_D = 0.7440$

Giá trị thống kê kê: $t_{qs} = \frac{\bar{x}_D}{s_D} \times \sqrt{n} = 2.3760$

Do $t_{qs} \in RR$ nên giả thiết H_0 bị bác bỏ.

Điểm trung bình của học sinh sau kỳ nghỉ đã giảm với mức ý nghĩa 5%.

- b) Mẫu X1: $n = 8$ $\bar{x} = 7.5$ $s = 1.1952$ $t_{0.005}(7) = 3.499$

Ngưỡng sai số của ước lượng: $\varepsilon = 3.499 \times \frac{1.1952}{\sqrt{8}} = 1.4786$

Khoảng tin cậy cần tìm: 7.5 ± 1.4786

c) Gọi p là tỉ lệ học sinh có tự ôn tập trong kỳ nghỉ.

$$Gt H_0 : p = \frac{1}{2}$$

$$Gt H_1 : p \neq \frac{1}{2}$$

$$Miền bác bỏ RR = (-\infty; -1.96) \cup (1.96; +\infty)$$

$$\text{Giá trị thống kê kê: } z_{qs} = \frac{f - p_0}{\sqrt{p_0(1-p_0)}} \sqrt{n} = \frac{\frac{40}{92} - 0.5}{\sqrt{0.5 \times 0.5}} \sqrt{92} = -1.2511$$

Do $z_{qs} \notin RR$ nên chưa bác bỏ được H_0 .

Có thể nói khoảng 50% học sinh đã tự ôn tập trong kỳ nghỉ.

Câu 3: (2 đ)

Gọi $\mu_1; \mu_2; \mu_3$ lần lượt là vận tốc đạp xe trung bình của 3 nhóm tham gia thực nghiệm.

$$Gt H_0 : \mu_1 = \mu_2 = \mu_3$$

$$Gt H_1 : \exists \mu_i \neq \mu_j \quad i \neq j$$

$$Miền bác bỏ RR = (3.89; +\infty)$$

$$x1_{tb} = 82.9 \quad x2_{tb} = 83.4 \quad x3_{tb} = 82.3 \quad x_{tb} = 82.8667$$

Ghi rõ các công thức tính....

ANOVA					
	Source of Variation	SS	df	MS	F
SSB	Between Groups	0.121333	2	0.060667	0.15463
SSW	Within Groups	4.708	12	0.392333	
SST	Total	4.829333	14		

Kết luận: Do $F_{qs} = 0.1546 \notin RR$ nên chưa bác bỏ được H_0 .

Chưa thể nói có sự khác biệt về hiệu quả của 3 chế độ dinh dưỡng tới thành tích thi đấu của các nhóm.

Câu 4: (0.5 đ + 0.75đ + 0.75đ)

Yêu cầu ghi rõ các công thức.

$$x_{tb} = 11.7143 \quad s_x = 3.6544 \quad y_{tb} = 8.7 \quad s_y = 2.6920$$

$$S_{XX} = 80.1286 \quad S_{XY} = 55.63 \quad S_{YY} = 43.48$$

a) $r = 0.9425$. Do $|r| > 0.8$ nên ta nói X,Y có quan hệ tuyến tính mạnh.

b) PT đường hồi quy tuyến tính mẫu: $\hat{y} = 0.5672 + 0.6943 x$.

$$\text{Dự đoán: } \hat{y}(10.5) = 7.8570$$

c) Cần ghi rõ từng bước tính. $SSR = 38.6216$

$$SSE = 4.8584$$

$$SST = 43.48 \dots$$

$$\text{Ngưỡng sai số } \varepsilon = \dots = 0.2831$$

$$\text{Khoảng tin cậy cho hệ số góc: } (0.4112; 0.9773)$$