2017년 2학기 확률과통계 기말고사

- 이름:
- 학번:

문제 1 (10점)

(1) 10마리의 말이 경주할 경우에 1, 2, 3 등을 맞출 확률을 계산하라.

(2) 10마리의 말이 경주할 경우에 3등 안에 들어갈 세 마리의 말을 맞출 확률을 계산하라.

힌트: 분수로만 계산한다.

(1) 10 마리 시합에서 1, 2, 3 등이 결정되는 경우의 수는
$_{10}P_3 = 10 \cdot 9 \cdot 8 = 720$ 이다.
따라서 1, 2, 3 등을 모두 맞출 확률은 $\frac{1}{720}$.

(2) (1) 문제에서 1~3 등의 순서가 중요하지 않다.
따라서 10 마리에서 세 마리를 순서없이 선택할 수 있는 경우의 수는
$_{10}C_3 = \frac{_{10}P_3}{3!} = 120$ 이다. 따라서 확률은 $\frac{1}{120}$.

문제 2 (5점)

주사위를 n번 던져서 2보다 큰 소수가 나오는 횟수의 분포는 이항분포를 따른다. 어떤 이항분포를 따르는지 설명하라.

1~6 숫자 중에서 2보다 큰 소수는 3과 5이다.
따라서 주사위를 한 번 던져서 2보다 큰 소수가
나올 확률은 $\frac{1}{3}$ 이다.
따라서 $B(n, \frac{1}{3})$의 이항분포를 따른다.

문제 3 (10점)

주어진 시간구간에 사건이 발생하는 수를 X라 하자. 만약 X가 동일한 시간구간마다 발생하는 횟수의 평균값이 λ이라면, X는 푸아송 분포를 따른다고 말하고 $X \sim Po(\lambda)$으로 표기한다.

(1) 푸아송 분포의 기대치와 분산을 구하라.

(2) X와 Y가 독립확률변수이고, 각각이 아래 모양의 푸아송 분포를 따른다: $X \sim Po(\lambda_x)$, $Y \sim Po(\lambda_y)$. 이때 $X + Y$는 어떤 분포를 따르는가? 또한 $X + Y$의 기대치 $E(X + Y)$와 분산 $Var(X + Y)$를 계산하라.

(1) $E(X) = Var(X) = \lambda$

(2) $X+Y \sim Po(\lambda_x + \lambda_y)$

$E(X+Y) = Var(X+Y) = \lambda_x + \lambda_y$

문제 4 (10점)

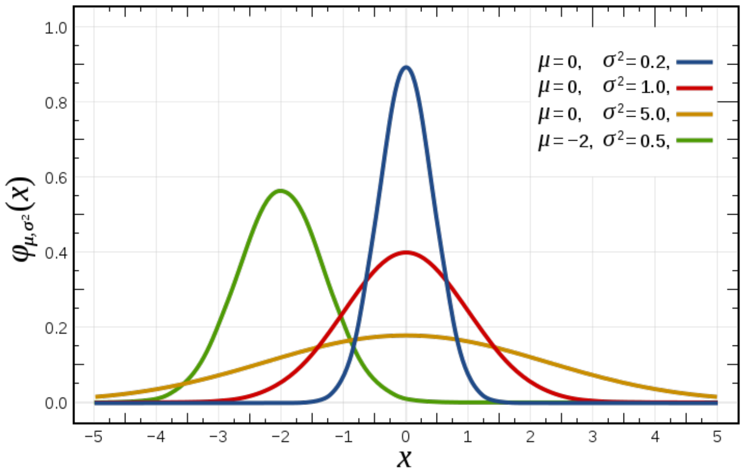
아래 데이터 중에서 이산데이터와 연속데이터를 구분하라.

이산 : 빨강색 동그라미 친 항목

1. 슬롯머신에서 정해진 상금액
2. 한국 성인 여성들의 키
3. 초등학교 학급당 인원 수
4. 올림픽 출전 선수들의 100m 달리기 기록
5. 동전던지기에서 앞면이 나온 횟수
6. 한라산에 있는 참나무 잎파리의 길이

문제 5 (5점)

아래 그래프는 세 개의 정규분포 확률밀도함수 곡선을 보여준다. 세 곡선의 공통점과 차이점을 설명하라.



μ는 평균 의미
σ^2는 분산 의미

① μ의 값에 따라 곡선의 중심위치가 좌우로 변경됨.

② σ^2의 값이 커질 수록 곡선의 높이가 낮고 완만해짐.

③ 모든 곡선에 대해, 곡선과 X축 사이의 면적은 1이다.

④ μ를 중심으로 좌우 대칭이다.

아래 표는 표준정규분포의 확률테이블이다.

	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990

문제 6 (10점)

$X \sim N(178, 5^2)$이라 하자. $P(X > 180)$와 $P(180 < X < 185)$을 계산하라.

$z_1 = 180$의 표준점수 $= \frac{180-178}{5} = 0.4$

$z_2 = 185$의 표준점수 $= \frac{185-178}{5} = 1.4$

$\Rightarrow$ $P(X>180) = P(Z>0.4) = 1 - P(Z \le 0.4) = 1 - 0.6554 = 0.3446$

· $P(180<X<185) = P(Z \le 1.4) - P(Z \le 0.4)$
$= 0.9192 - 0.6554 = 0.2638$

문제 7 (10점)

이것은 사실상 $P(X=8)$의 값을 요구하는 것이다. 그리고 실제로 $P(X=8) = {}_{16}C_8 \cdot (\frac{1}{2})^{16} = 0.1964$이다.

근사치임

$X \sim B(16, \frac{1}{2})$라고 하자. 정규분포 및 연속성 보정을 활용하여 아래 질문에 답하라.

(1) $P(7 < X)$을 계산하라.

(2) $P(7 < X < 8.5)$을 계산하라.

$16 \cdot \frac{1}{2} = 8 > 5$ 이므로 $B(16, \frac{1}{2})$은 근사적으로 정규분포 $N(8, 2^2)$를 따른다.

(1) $P(7<X)$
$= P(7.5 < X)$ ← 연속성 보정
$= P(\frac{7.5-8}{2} < Z)$ ← 표준화
$= P(-0.25 < Z)$
$= P(Z \le 0.25)$
$= 0.5987$

(2) $P(7<X<8.5)$ X는 이산자료임
$= P(7<X<9)$
$= P(7.5<X<8.5)$
$= P(-0.25<Z<0.25)$
$= P(Z<0.25) - P(Z<-0.25)$
$= P(Z<0.25) - (1 - P(Z<0.25))$
$= 2 \cdot P(Z<0.25) - 1$
$= 0.1974$

중심극한정리

모집단 X에서 표본을 추출할 때 표본의 크기 n이 충분히 크면 $\bar{X}$의 분포가 개략적으로 정규분포를 따른다.

또한 $E(X) = \mu$이고 $Var(X) = \sigma^2$이면 다음이 성립한다.

$$\bar{X} \sim N(\mu, \sigma^2/n)$$

문제 8 (10점)

영화관의 팝콘기계의 주당 평균 고장횟수는 푸아송 분포를 따르며 $X \sim Po(3.4)$가 성립한다. 30주 동안 조사해 보았을 때 주당 고장횟수의 평균이 4회 이상일 확률을 구하라.

힌트: $\sqrt{0.113} = 0.337$

$X \sim P_o(3.4)$ 이므로 $\bar{X} \sim N(3.4, \underbrace{0.113}_{\frac{3.4}{30}})$ 이다.

따라서 30주 동안 4회이상 고장날 확률은 다음과 같다.

$$P(\bar{X} \geq 4) = P\left(Z \geq \frac{4-3.4}{\sqrt{0.113}}\right)$$
$$= P\left(Z \geq \frac{0.6}{0.337}\right) = P(Z \geq 1.78)$$
$$= 1 - P(Z \leq 1.78) = 1 - 0.9625$$
$$= 0.0375$$

아래 표는 신뢰구간 설정 공식을 경우별로 담고 있다.

모집단 통계	모집단 분포	조건	신뢰구간
μ	X 정규분포	σ^2 알고 있음 n 표본크기 $\bar{X}$ 표본 평균값	$\left(\bar{X} - c\frac{\sigma}{\sqrt{n}}, \bar{X} + c\frac{\sigma}{\sqrt{n}}\right)$
μ	X 정규분포 아닌	σ^2 알고 있음 $n \geq 30$ $\bar{X}$ 표본 평균값	$\left(\bar{X} - c\frac{\sigma}{\sqrt{n}}, \bar{X} + c\frac{\sigma}{\sqrt{n}}\right)$
μ	X 임의의 분포	σ^2 모름 $n \geq 30$ $\bar{X}$ 표본 평균값 s^2 표본 분산	$\left(\bar{X} - c\frac{s}{\sqrt{n}}, \bar{X} + c\frac{s}{\sqrt{n}}\right)$
P	이항	$n \geq 30$ P_s 표본비율 $q_s = 1 - P_s$	$\left(P_s - \underbrace{c \cdot \sqrt{\frac{P_s \cdot q_s}{n}}}_{\text{오차범위}}, P + c \cdot \sqrt{\frac{P_s \cdot q_s}{n}}\right)$

오차범위는 기본적으로 아래의 형태이다:

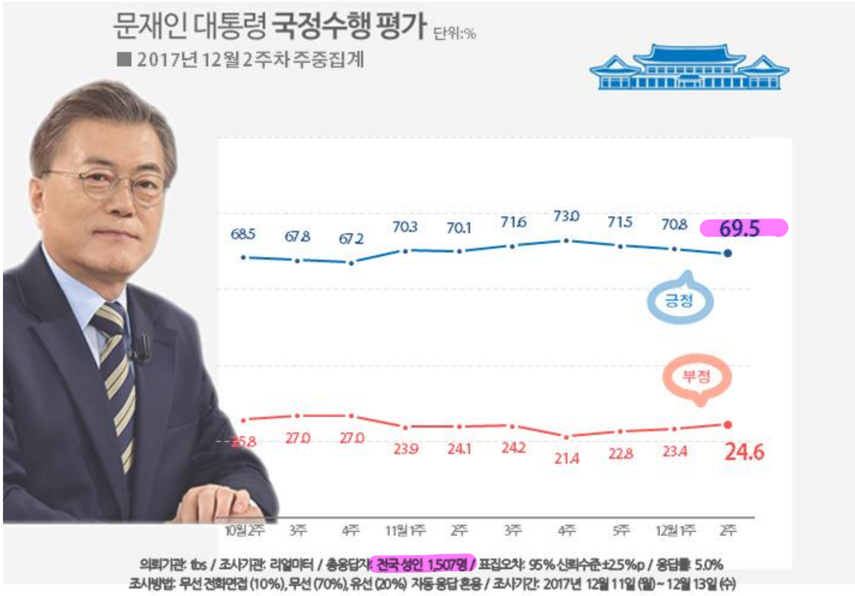
$\boxed{C * (\text{표본 표준편차})}$

* C 의 값 : 신뢰수준에 의해 결정됨

신뢰수준	C 의 값
90 %	1.64
95 %	1.96
99%	2.58

문제 9 (10점)

지지율 설문조사가 대표적인 표본비율의 분포를 사용한다. 아래 그림에서 언급된 표집오차(95% 신뢰수준 ± 2.5%p)를 설명하라. (힌트: 정확한 계산은 요구하지 않음)



신뢰수준 95% : 동일한 조건에서 설문조사할 경우 100의 95번은 신뢰오차 범위 내에서 비슷한 결과를 얻을 것이라는 의미

± 2.5 p : 신뢰구간을 정할 때 사용하는 신뢰오차이며 아래 식에 의해 결정된다.

$$\pm C \cdot \sqrt{\frac{p_s \cdot q_s}{n}}, \quad \text{단} \begin{cases} C = 1.96, \\ n = 1507 \\ p_s = 0.695 \\ q_s = 1 - p_s \end{cases}$$

문제 10 (10점)

가설검정 6단계를 순서별로 나열하라.

3 → 6 → 1 → 4 → 5 → 2

1. 기각역 설정
2. 가설 수용여부 판단
3. 영가설과 대립가설 설정
4. p-값 확인
5. p-값의 기각역 포함 여부 확인
6. 검정통계 설정

문제 11 (10점)

어느 고등학교에서 지난 몇 년 동안 학교생활만족도를 조사한 결과 만족도 평균값이 78%였다. 올 해 조사한 만족도는 70%였다. 학교생활만족도가 의미있게 변화하였는지 여부를 가설검정하고자 할 때 필요한 영가설과 대립가설을 설정하라.

p 를 만족도라 하자.

영가설 (H_0) : $p = 78$

대립가설 (H_1) : $p < 78$

와지만

H_1 : $p \neq 78$ 도 인정.

이유: "의미있게 변화하였는지 여부" 라는 표현에서 "의미있게" 라는 것은 만족도가 급격히 떨어진 것 라 할게 급격히 상승한 것도 내포할 수 있음.