

1. 다음은 “인터넷 댓글문화에서 플레이밍 행동에 영향을 미치는 요인에 관한 연구”의 논문의 일부를 발췌한 내용이다.

플레이밍(flaming)은 컴퓨터 환경에서 특정의 사용자들이 무례함이나 감정적 행동을 분출하는 것으로(Spitzer 1986), 타인에게 단순히 화를 내는 것뿐만 아니라 적대감과 같은 강한 감정을 분출하는 것에 이르기까지 반사회적이고 부정적인 행동을 포괄하는 개념이다(Lea, O’Shea, Fund, and Spears 1992; Suh and Wagner 2013).

연구문제 1: 온라인 상에서 플레이밍 행동에 영향을 미치는 상황적, 개인적, 동기 요인들은 무엇인가?

플레이밍 행동 경험은 이철선 (2003)이 사용한 척도를 이용하여 “무례한 언어를 사용한 적이 있다”, “타인을 무시하는 태도를 보인 적이 있다”, “타인에게 적대적인 태도를 표현한 적이 있다”를 전혀 “그렇지 않다”에서 “매우 그렇다”의 5점 리커르트 척도로 구성하였다. 플레이밍 행동 의도는 이철선 (2003)이 사용한 척도를 이용하여 “나는 상대방의 의견을 비웃는 댓글을 달 것이다”, “나는 상대방에게 적대감을 표현하는 댓글을 달 것이다”, “나는 상대방의 의견을 무시하는 댓글을 달 것이다”를 전혀 “그렇지 않다”에서 “매우 그렇다”의 5점 리커르트 척도로 구성하였다

(1) sample space를 하나 정의하고, 정의된 sample space 상에서 random variable을 하나 정의하라.

(2) 이 연구에서 사용할 수 있는 statistic을 하나 정의하여 보아라.

2. 최빈치, 중간값, 중앙값은 어떤 경우에 가장 적절한 대표치인지에 대하여 생각하여 보라.

3. 평균과 분산은 어떤 관계가 있는지를 손실함수(loss function)의 관점에서 설명하라.

4. 조건확률과 Bayes 정리를 설명하라. (Bayes정리의 경우는 공식을 외우지 못하면, 어떤 경우에 사용하는지 어떤 특징이 있는지를 기술하면 됨)

5. “동전을 던졌을 때 앞면이 나올 확률이 1/2이다”의 의미에 대하여 설명하라 (확률의 의미를 중심으로 설명하라)

6. 정규분포의 확률밀도함수를 외어서 쓰라.

7. 중앙극한치정리 (Central Limit Theorem)를 기술하라.

8. 확률분포함수와 확률밀도함수는 어떤 점에서 다른가?

9. 가설검증에서 일종오류(Type I error), 이종오류(Type II error),  $\alpha, \beta$ , power의 관계에 대하여 설명하라.

10.  $\sqrt{\frac{\sum(X_i - \bar{X})^2}{N}}$  과  $\sqrt{\frac{\sum(X_i - \bar{X})^2}{N-1}}$  의 차이를 설명하라.

11.  $R_1, R_2, R_3, \dots, R_{10} \stackrel{iid}{\sim} B(10, 0.7)$  일때,  $P(R_1 = 5 \cap R_2 = 6)$ 를 구하라.

[다음의 정보를 이용하라]

iid: independently identically distributed

$$P(R_5 = 5) = 0.10$$

$$P(R_6 = 6) = 0.20$$

12. 표본분산의 표집분포 (sampling distribution of sample variances)은 어떤 의미인지를 설명하라.

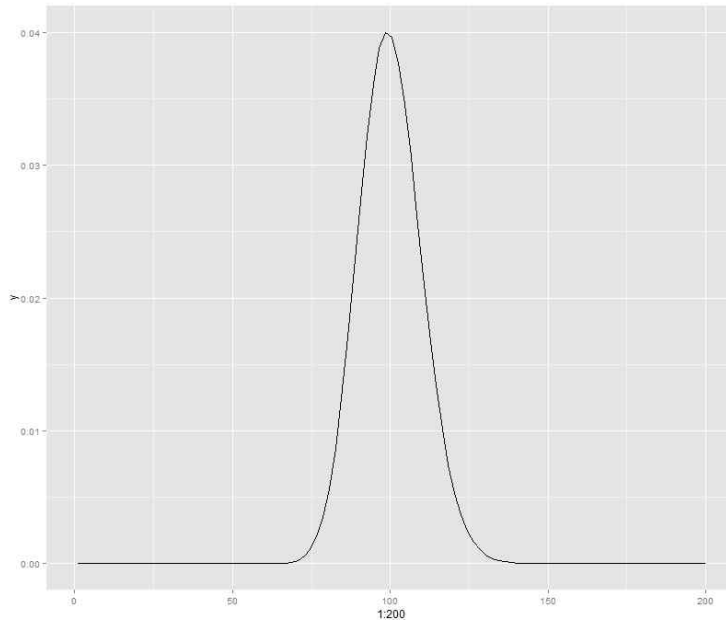
13. 다음의 (1)-(6)의 빈칸을 채워라. 모든 자료는 iid인 정규분포를 따른다고 가정하라. Test Statistic에서 분모의 경우는 하나의 계산을 위한 완벽한 공식을 제시할 필요는 없으며 어떤 변수의 표준편차인지를 나타낼 수 있으면 된다.

영가설	$\sigma$	자료	Test Statistic	Test Statistic의 분포
$\mu=100$	known	하나의 자료 (X)	$\frac{X-\mu}{\sigma_x}$	정규분포
$\mu=100$	unknown	sample mean ( $\bar{X}$ )	(1)	(2)
$\mu_1=\mu_2$	unknown	sample mean and sample standard deviation	(3)	(4)
$\mu_1=\mu_2$	unknown	동일한 사람에게서 나온 두 측정치의 차이의 평균 ( $\bar{D}$ )	(5)	(6)

14. 각각 100명으로 구성된 남녀 표본의 표본평균을 계산해보니, 110과 120이었다. 이 두 표본평균이 다르다는 것을 보여주기 위하여 어떤 통계적 절차를 거쳐야 되는가? [read very carefully. It is a tricky question]

15. 당신은 작은 가게를 운영하고 있는데 최근 고객의 수가 늘어난 것 같아서, 이를 통계적으로 검증하려고 한다. 그동안 1분에 평균 1명의 고객이 방문하고 있었다. 오늘 100명의 고객이 방문할 때까지 시간을 측정하니 80분이 소요되었다. 오늘의 고객의 수가 통계적으로 유의하게 늘었는지를 검증하려한다. 다음의 사실을 고려하라.

- 1분에 평균  $\beta$ 명의 고객이 방문할 때, 100명의 고객이 방문할 때까지 걸리는 시간을 무선변수  $X$ 로 잡았을 때,  $X$ 는 gamma 분포를 따르고 파라미터는  $\beta$ 이다. 아래의 그래프는  $\beta$ 가 1인 gamma 분포의 확률밀도함수이다.



(1) 연구의 목적에 맞게 영가설을 어떻게 설정해보라.

(2) 이 영가설에 맞추어서 가설검증을 하여보라.

위의 감마분포에서 계산된 확률은 다음과 같다.

$$P(X < 80) = 0.017$$

$$P(X < 100) = 0.513$$

[Gamma 분포를 배우지 않았지만, 가설검증을 이해하고 그 논리를 따르면 쉽게 풀 수 있는 문제임]