**객관적자료분석 HW#1**

지구환경과학부

2017-29008 백관구

**1.1.** 주어진 예제의 95% 유의수준에서 분산의 신뢰구간을 추정하기 위하여 검정을 수행한다.

자유도 , 표준편차 , , 이므로 주어진 예제의 **95%** 유의수준에서 분산은 신뢰구간에 존재한다.

**1.2.** 만일 주어진 예제가 정규분포라고 착각하여 S를 로 잘못 두고 z 검정을 수행할 경우 95% 유의수준에서 실제 평균값은 일 때,

따라서 **95%** 유의수준에서 추정한 평균의 신뢰구간은 로 t 검정으로 추정한 평균의 신뢰구간보다 좁은 범위로 추정하게 된다.

**1.3.** 가설 검증은 5 단계를 거치게 된다.

1. 유의수준: 유의수준 95%에서 유의성 확인한다.
2. 귀무가설: 화산폭발 한 달 전 10 개 화산의 월평균 아노말리 기온과 세 달 후 월평균 아노말리 기온의 차이가 없다 ().
3. 샘플의 수 으로 이므로 t 검정을 수행한다.

화산의 개수 , 자유도 , , 월평균 아노말리 기온의 표준편차 이므로,

1. 화산폭발 전후 월평균 아노말리 기온의 차이 이므로 귀무가설을 채택한다. 따라서 유의수준 **95%** 내에서 화산폭발 전후의 월평균 아노말리 기온은 **유의미한 차이를 가지지 않는다.**

**1.4.** 샘플의 수 으로 충분치 않으므로 유의수준 99%에서 t 검정을 수행한다.

평균 적설량 , 자유도 , , 표준편차 이므로 유의수준 **99%**에서 **t 검정**을 통해 추정한 신뢰구간은 이다.

**1.5.** 두 집단 각각의 평균이 유의미한 차이를 보이는지 검증한다.

1. 유의수준: 유의수준 95%에서 유의성을 확인한다.
2. 귀무가설: 두 집단 각각의 평균이 차이가 없다 ().
3. 샘플의 개수 , 로 이므로 t 검정을 수행한다.

두 집단 각각의 표준편차 , 이고 자유도는 이므로 , 이다.

1. 두 집단 평균의 차이 이므로 귀무가설을 기각한다. 따라서 유의수준 **95%** 내에서 **전 기간 대비 1920~30년대의 강수량은 유의미한 차이**를 보인다.

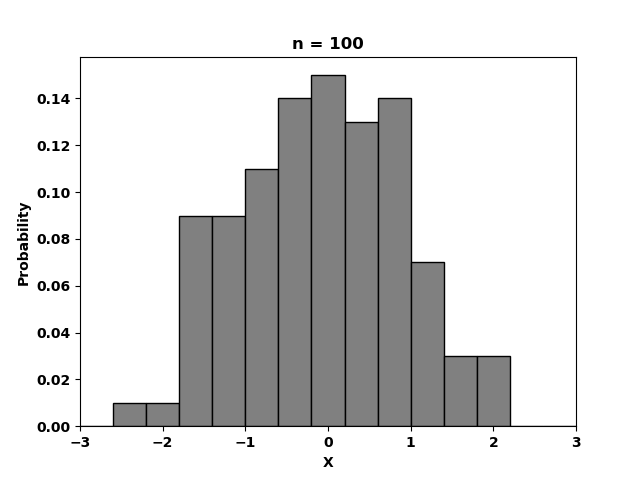
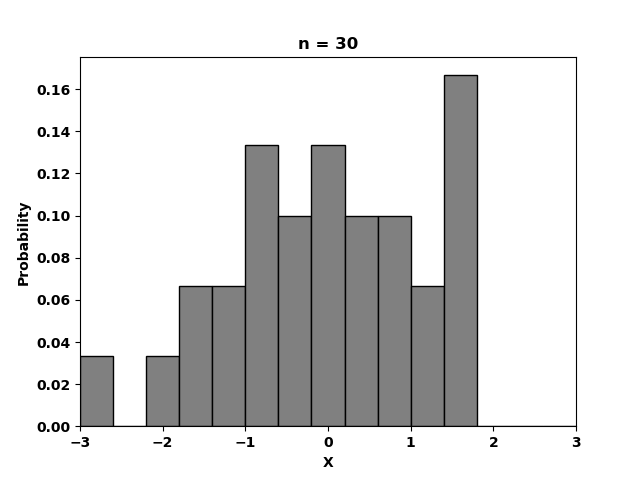
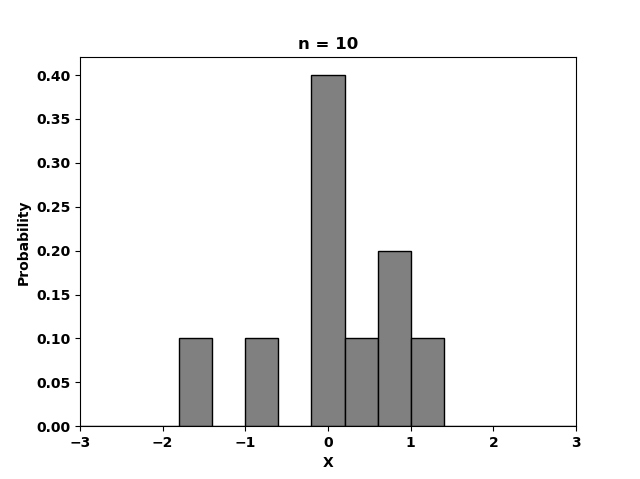
**1.6.** 정규분포 N(0, 1)를 갖는 난수를 10, 30, 100 개를 생성하여,

1. 평균 , 분산 , 왜도 , 첨도 를 계산

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **n=10** | **n=30** | **n=100** | **n=1000** |
| **평균** | **0.0598** | **-0.0274** | **-0.1042** | **-0.0105** |
| **분산** | **0.5655** | **1.4198** | **0.9404** | **0.9990** |
| **왜도** | **-0.4937** | **-0.4099** | **-0.0178** | **-0.0689** |
| **첨도** | **3.1006** | **2.4253** | **2.3700** | **2.9899** |

이론상으로 샘플의 개수가 많아질수록 평균은 0, 분산은 1, 왜도는 0, 첨도는 3에 가까워져야 한다. 분산과 왜도는 샘플의 수를 10🡪30🡪100으로 늘릴수록 그러한 경향을 보였다. 하지만 정규분포 N(0, 1)을 갖는 난수 생성을 하였음에도 불구하고 평균은 n=100일 때 가장 0에서 멀었고 첨도도 3에서 가장 먼 값을 보였다. 그래서 추가적으로 n=1000일 때 같은 계산을 한 결과, 정규분포에 가까운 통계치을 얻을 수 있었다.

1. 확률분포함수



샘플의 분포를 살펴보면 정규분포 형태에 점차 근접하는 것을 알 수 있다. n=10일 때 음의 값으로 꼬리가 길게 나타나 왜도가 작은 음수를 가졌고, 0 부근에 많이 분포해 첨도가 높게 나타났다. n=30일 때도 음의 값 쪽으로 꼬리가 길어 왜도가 음의 값을 가졌다. n=100일 때는 정규분포에 가장 가까운 분포 형태를 보였다.

1. 유의수준 95%에서 , 의 신뢰구간 추정
2. 신뢰구간 추정

n=10, n=30일 때 샘플이 충분치 않으므로 t 검정을 수행하고, n=100일 때 z 검정을 수행한다.

|  |  |  |
| --- | --- | --- |
| **n=10** | **n=30** | **n=100** |
|  | |  |
|  |  |  |

1. 신뢰구간 추정

분산의 신뢰구간을 추정하기 위하여 검정을 수행한다.

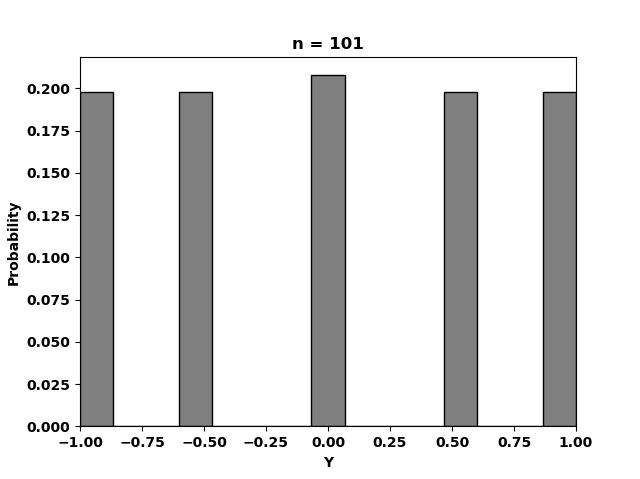
|  |  |  |
| --- | --- | --- |
| **n=10** | **n=30** | **n=100** |
|  | | |
|  |  |  |

같은 유의수준이라도 샘플의 개수가 많아질수록 평균과 분산의 신뢰구간이 좁혀지는 것을 알 수 있다. 즉, 표본이 많을수록 모평균, 모분산에 대한 추정이 보다 정확해진다.

**1.7.** 에 대해서 평균, 분산, 왜도, 첨도를 계산, 확률분포함수 표현

|  |  |  |  |
| --- | --- | --- | --- |
| **평균** | **분산** | **왜도** | **첨도** |
| **0.0000** | **0.4950** | **0.0000** | **1.5150** |

-1과 1 사이에서 값이 반복되는 주기함수이므로 평균과 왜도는 0이 나온다. 반면 분산과 첨도는 정규분포(분산 1, 첨도 3)라 지칭할 수 있는 값과 멀리 떨어져 있는 것으로 보인다.



표본의 분포가 변곡점과 최대 및 최소값에 밀집되어 나타난다.