

授業資料

1. スライドにページ番号をつけていただくと復習しやすいです。

了解です。今後はページ番号を付けるようにしておきます。

2. 配布資料において、前回の出席票に対する回答の文字が小さくて見づらいです。もう少し大きくできないでしょうか。

了解です。PowerPoint スライドにするのは無理があるようなので、通常のテキストとして読めるようになったはずですが、いかがでしょうか。

3. 前回課題の返却方法は、大変効率が悪いので、レポート BOX での返却に戻していただけますか？ 現状でも評価は人に見られているので、プライバシー云々はどちらにしる同じだと思うのですが。

レポート返却 BOX を使うとどの程度返却されたか把握できないという問題も感じたので、通常課題に関しては、授業にあわせて返していくことにしています。ただし、中間レポート及び最終試験の分は、レポート返却 BOX を用いたいと思います。できるだけ返却は効率的にできるようにしますので、協力をお願いします。

中間レポート課題

4. 冬休みの課題で参照元の「写し」とは何か、が良く分からなかったなので、もう一度具体的に何を書き記すか教えてください。

5. レポートには写しも付けること、とありますが、どういう意味でしょうか？

例えば、新聞記事や雑誌記事でしたら、その内容が分かるよう記事全体を A4 版に複写コピーしたものを添付してください（コピー分はレポートの枚数制限を超過しても構いません）。

6. 中間レポートは、両面刷りと片面刷りのどちらがよいでしょうか。

今回の中間レポートに関しては、片面でお願いします。

7. 中間レポートでも手書き可ですか？

8. 中間レポートも手書きで大丈夫なのでしょうか？

9. 中間レポートは、手書きとハードコピーのどちらでもよいのでしょうか。

これまでの課題と同様、手書きでもデジタル作成でもどちらでも構いません（手書きの人はできるだけ丁寧な字にしましょう）。

10. 中間レポートの演習問題で使うデータは、でっあげで良いですか？ また、データ数はどのくらいあるのが望ましいでしょうか。

データそのものは仮想的なもので構いません（もちろん、実データであってもよいです）。またデータ件数も特に指定しません。適宜、その設問にあうと思われる題材・件数を使ってください。

評価

11. 授業に出て無い人が不正をして出席扱いになっているので、不安です。

12. 友人に授業前に前回レポートの回収と出席票を頼んで、2限の前に来て、レポートだけ出したり、代返のようなことをしている人がいます。レポートも出席票も出している場合は判別できないけど、レポートの回収だけして、出席票出していない人は明らかに代返ではないでしょうか？ 正直、そういう人にはペナルティを与えて欲しいと思っています。

対応のため、年明けの授業分からは出席点をカウントしないこととします。

13. テスト 7割、課題 3割ということは、出席点は加算されないと考えてよいですか？

出席点は 1割分（おまけの一種）として加算しています。

14. 今回まで、これからの課題の模範解答を教えてくださいませんか。

了解です。年明けまでに順次 Moodle にあげておきます。

15. 長谷川先生は計算ミスでの減点はあまりしなかったと思うのですが。

この授業は長谷川先生の授業ではありませんので悪しからず。

標本分散・母分散

16. E は期待値だけでなく、平均も表すのですか？

17. なぜ期待値が平均値といえるのですか？

「期待値」と「平均値」は同一のものです。それぞれの定義式から明らかです（教科書 p.53 も参照のこと）。確率の話題のときは「期待値」と呼ぶことが多いようですが、両者は可換ですので、適宜読み替えられるようにしてください。

18. 標本分散の平均と、標本平均の分散の違いがよく分かりません。

19. $V(\bar{x}) = \frac{\sigma^2}{n} = S^2$ ということですか？

整理しておかないと結構ややこしいところだと思うので、再度復習したいと思います（スライド参照）。

20. テキストの p.93 で $E(X)$ と $V(X)$ はどうして回数と確率をこのように計算したら出てくるのか、いまいち理解できなかったのですが、どのような仕組みですか？

テキスト p.74-75 を参照ください。二項分布においては、その事象が発生する 0 と 1 が、試行回数分を繰り返されることが毎回独立に生じることがその定義ですから、その事象が発生する確率 p が試行回数 n 回分繰り返された分だけが常に発生することを期待できるはずですが、例えば、サイコロを振って 6 の目が出る確率は $\frac{1}{6}$ ですが、1 回振ったときの期待値は当然 $\frac{1}{6}$ となり、2 回振ったときは $\frac{2}{6}$ となり、...、6 回振ったときには $\frac{6}{6} = 1$ となるはずですが。これは試行回数が何回に増えても同じ性質を持つことが保証できるので、 $E(X) = np$ として考えることができます。

21. 分散を求める際、 n で割るべき？ $n-1$ で割るべき？ 違いの理由が知りたいです。

「分散」という語を使う場合は n で割るべきと思います。ただし、「標本分散」は

$$\frac{1}{n} \sum_{k=1}^n (X_k - \bar{X})^2 \quad (1)$$

$$\frac{1}{n-1} \sum_{k=1}^n (X_k - \bar{X})^2 \quad (2)$$

といった式で表される流儀がそれぞれあるようです。文献や解析ソフトウェアにより呼び方が異なる場合がありますので、注意が必要です。

この講義（及び教科書）では、式 (1) を分散または標本分散とし、式 (2) は「標本の不偏分散」と呼び、明確に区別して用います。標本の不偏分散は今日の講義で出てくる予定です。

22. 半整数補正は n がどの程度まで行えばよいのでしょうか？

基本的に、二項分布は離散的な分布となっているため、相当に大きな n （例えば 10,000 等）を取らない限り、半整数補正した方が近似が効いた値となります。

23. 「十分に大きい」という表現が頻出しますが、そう判断する基準はありますか？ 具体例（実験試行なら x 回、対人調査なら y 人）などをお願いしたいです。

2 つほど例を挙げる。

まず、標本抽出における復元・非復元抽出の際の違いについては、大まかな目安として抽出率 $(\frac{n}{N})$ が 5% 以下ならば、母集団が十分に大きい（もしくは、標本サイズが小さい）ので、復元・非復元の抽出方法の違いは無視できると言える。例えば、テキスト p.100 の例にある世論調査では $\frac{5000}{108470000} = 0.0046\%$ 、新入生の身長では $\frac{900}{35000} = 2.57\%$ なので、復元抽出としてみなして分布を考えてよい。

また、「中心極限定理」においては、「標本サイズ n が十分に大きいときは標本平均 \bar{X} は近似的に正規分布 $N(\mu, \frac{\sigma^2}{n})$ に従う」（テキスト p.107）と定義されているが、この場合の標本サイズ n は 100 程度のサイズになれば安定した近似ができるようになることが知られている（近似すべき対象・有効数字にも依る）。

24. 乱数表の使い方において、必要に応じて適宜「200, 400, 800 を引いて、すべてを 000~999 に収める」（p.103）とありますが、引く数が任意であるならば、人によって、2桁の数が多くなるとか、癖が出てしまうのではないのでしょうか？ 意識してバラバラにするしかないのでしょうか？

はい、恣意的に選んだ乱数列の選択によって、

小さな値が出やすいといった形で偏りが出てしまうことは大きな問題となります。ただし、テキストにおける例（引く数による操作を加える場合）では、とある一定の数を用いるという点での恣意性を加えていますが、その癖も含めて乱数列全体にわたって偏りが出る形にはなっていないので問題とはなっていません。

25. 教科書の p.108 では、標本抽出のために復元抽出を行っていますが、この場合、非復元抽出を行ってはいけないのですか？

非復元抽出を行っても構いませんが、復元抽出と非復元抽出では、出てくる統計量が全く異なる値になることに、注意してください。

26. なぜ乱数サイは正 20 面体なのでしょう。正 10 面体でもよいのではないのでしょうか？

「正 10 面体」というものは作れないことが知られていますので、正 20 面体になっています。

その他

27. $\sqrt{2} = 1.41 \approx 1.4$ となることがあるのですが、最初の等号が \approx でないのはなぜでしょうか？

最初の等号も無理数を丸めた近似値に直している点では変わりありません。指摘の通り、両者の等号は \approx と書くことで問題ありません。

28. 前の奴がずっと東方やっててうっとうしい。周囲の集中力を削がないよう、配慮をお願いします。