Praxisbeispiel zur Selbstlerneinheit: Prüfungsergebnisse analysieren

Autor*in: Andrea Ortiz

Ziele der Analysen

- Evaluation der Prüfungsergebnisse „Grundlagen des Verstärkungslernens“ aus zwei Semestern, um Rückschlüsse zur Verbesserung der Prüfung ziehen zu können
- Entwicklung eines (möglicherweise MATLAB-basierten) Tools zur systematischen Analyse der Ergebnisse künftiger Prüfungen

Lehrveranstaltung

Die Veranstaltung „Fundamentals of Reinforcement Learning“ mit zwei Vorlesungen und einer Übung des Fachgebiets Kommunikationstechnik am Fachbereich Elektro- und Informationstechnik, ist ein Wahlfach im Master verschiedener vom Fachbereich angebotener Studiengänge. Die Vorlesung wird jedes Sommersemester angeboten und begann im Sommersemester 2021. Obwohl die Zahl der angemeldeten Studierenden in der Regel groß ist – ca. 100 –, gibt es normalerweise ca. 30 zu Prüfende. Bei den Wiederholungsprüfungen im Winter ist die Zahl der Studierenden noch einmal geringer. Daher wurden für die Analyse nur die Prüfungen des Sommersemesters betrachtet.

Vergleich der Prüfungen

Beide Prüfungen – SoSe 2021 und SoSe 2022 – haben dieselbe Struktur: eine Aufgabe zur Bewertung der grundlegenden Konzepte und vier Rechenaufgaben. Außerdem behandeln diese vier Aufgaben in beiden Prüfungen die gleichen Themen. Die Höchstpunktzahl für die Prüfungen beträgt 60 Punkte. Sie gilt als bestanden, wenn die Studierenden mehr als 20 Punkte erreichen.

Abbildung 1 zeigt die durchschnittliche Punktzahl pro Studierenden in den beiden Prüfungen. Trotz des geringen Rückgangs der durchschnittlichen Punktzahl in SoSe 2022, ist die Standardabweichung im Vergleich zu SoSe 2021 geringer. Das bedeutet, dass der Unterschied zwischen den Noten der Studierenden kleiner ist. Möglicherweise könnte dies darin begründet sein, dass die Prüfung im SoSe 2021 zum Ersten Mal durchgeführt wurde und die Studierenden eventuell schlechter abschätzen konnten, wie die Prüfung aussieht. In späteren Semestern lagen in der Studierendenschaft bereits Beispiele für den Inhalt der Prüfung vor und eine bessere Vorbereitung war somit möglich.

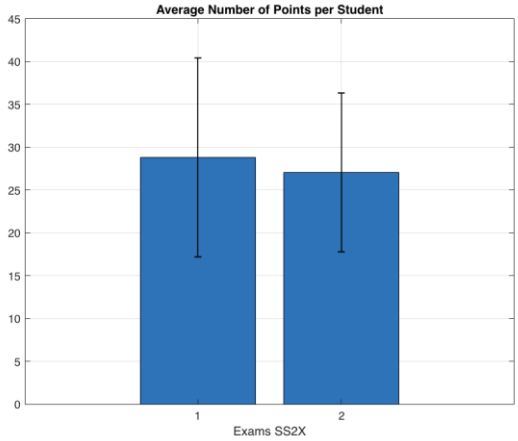


Abbildung 1: Durchschnittliche Punktzahl pro Studierenden in den beiden Prüfungen

Die Histogramme der Verteilungen der Punkte pro Aufgabe sind in den Abbildungen 2 und 3 für das SoSe 2021 bzw. SoSe 2022 dargestellt. Die roten, gelben und grünen Linien zeigen die minimale, durchschnittliche und maximale Anzahl der erreichten Punkte.

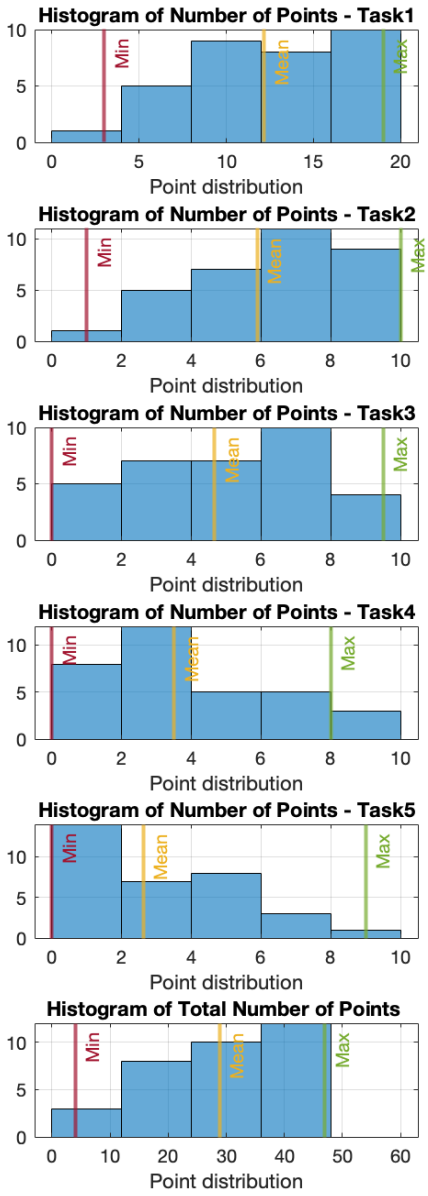


Abbildung 2: Histogramme der Verteilungen der Punkte pro Aufgabe – SoSe 2021

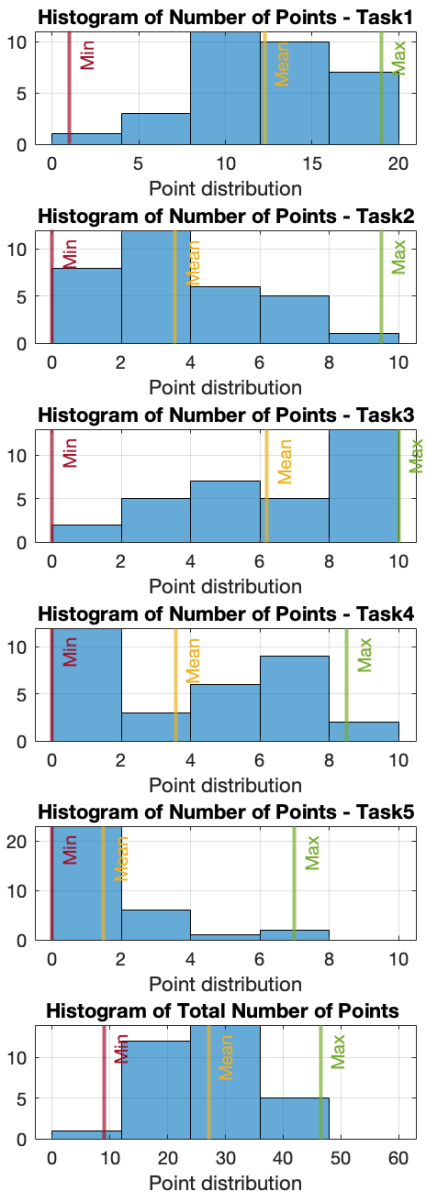


Abbildung 3: Histogramme der Verteilungen der Punkte pro Aufgabe – SoSe 2022

Bei **Aufgabe 1** (Task1) sind die durchschnittliche und die maximale Punktzahl ähnlich. Allerdings war im SoSe 2021 die minimale Anzahl der erhaltenen Punkte größer als im SoSe 2022.

Bei **Aufgabe 2** (Task2) ist das Verhalten in beiden Semestern entgegengesetzt. Im SoSe 2021 waren mehr Studierende in der Lage, größere Teile der Aufgabe zu lösen, und erreichten folglich mehr Punkte. Im SoSe 2022 bezog sich die Aufgabe zwar auf dasselbe Thema, aber auf ein anderes Niveau in der Blooms Taxonomie.

In **Aufgabe 3** ist ein besseres Durchschnittsverhalten (durchschnittliche Punktzahl) für SoSe 2022 zu beobachten. Im Vergleich zu SoSe 2021 war die Komplexität von Aufgabe 3 in SoSe 2022 geringer.

Die **Aufgaben 4 und 5** weisen in beiden Semestern in etwa das gleiche Verhalten auf. Insbesondere ist zu erkennen, dass die Studierenden bei Aufgabe 5 im Durchschnitt eine

niedrige Punktzahl erreichten. Der Grund für diesen Trend könnte sein, dass Aufgabe 5 die letzte Aufgabe der Prüfung ist. Wie einige Studierende berichteten, hatten sie kaum Zeit, um sie zu lösen, obwohl sie unabhängig von den vorherigen Aufgaben ist.

Die Abbildungen 4 und 5 vergleichen die geschätzten Aufgabenschwierigkeiten in den beiden betrachteten Semestern. Es ist zu erkennen, dass keine der Aufgaben in die Kategorie „leicht“ fällt, sondern im Bereich der akzeptablen Schwierigkeit liegt. Tatsächlich wird im SoSe 2022 nur Aufgabe 5 als schwierig eingestuft. Von der Konzeption her ist die Aufgabe nicht schwieriger als im SoSe 2021. Die Metrik berücksichtigt jedoch die Leistung der Studierenden, um die Schwierigkeit einzuschätzen. Wie bereits erwähnt, haben aus Zeitgründen weniger Studierende die Aufgabe 5 gelöst und daher auch weniger Punkte erreicht.

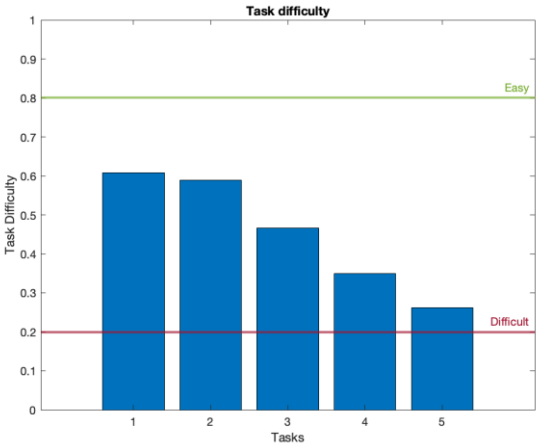


Abbildung 4. Aufgabe Schwierigkeit SoSe 2021

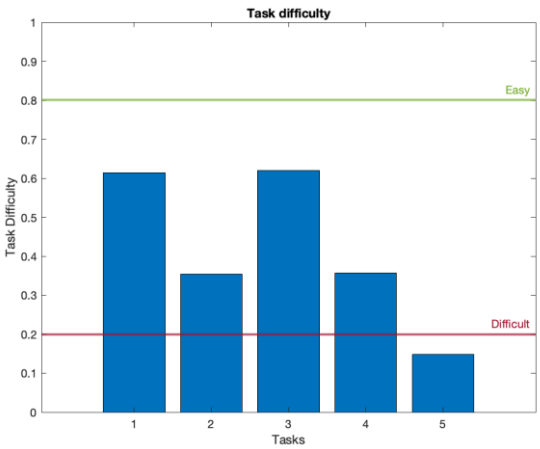


Abbildung 5. Aufgabe Schwierigkeit SoSe 2022

Die Trennschärfe der Aufgaben in den beiden Semestern ist in den Abbildungen 6 und 7 dargestellt. Obwohl keine der Aufgaben unter den akzeptablen Grenzwert fällt, hatte die Aufgabe im SoSe 2021 eine bessere Trennschärfe als die in SoSe 2022. Die Interpretation dieser Metrik ist, dass im SoSe 2021 Studierende, die eine hohe Punktzahl erreichten, durchweg gute Leistungen in allen Aufgaben zeigten. Im Gegensatz dazu war in SoSe 2022 die Verteilung der Punkte vielfältiger. Das bedeutet, dass Studierende mit einer insgesamt niedrigen Leistung in der Lage waren, in einer bestimmten Aufgabe gut zu sein, insbesondere in Aufgabe 3, die den niedrigsten Wert hat.

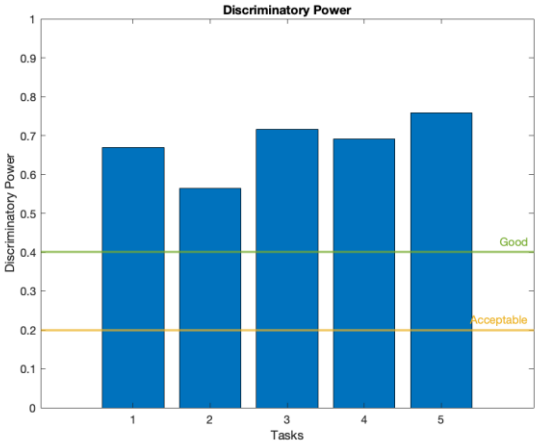


Abbildung 6. Trennschärfe SoSe 2021

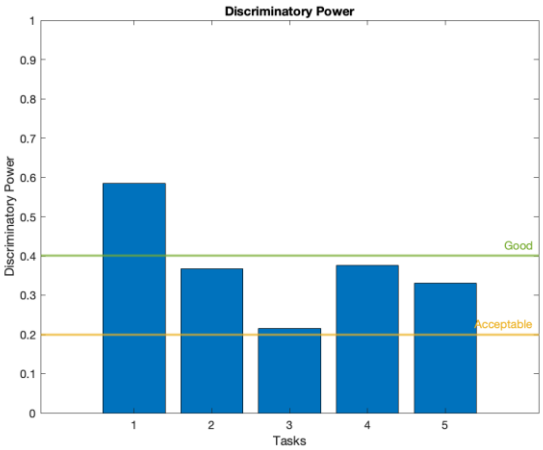


Abbildung 7. Trennschärfe SoSe 2022

Mit dem Vergleich von **Cronbachs Alpha** in den Abbildungen 8 und 9 kann die Reliabilität der Aufgaben bewertet werden. In der Abbildung zeigt die gelbe Linie das Gesamt-Cronbachs Alpha. Die Balken im Balkendiagramm geben das Cronbachs Alpha an, wenn jede der Aufgaben weggelassen wird. Die Ergebnisse von SoSe 2021 zeigen ein größeres Cronbachs Alpha im Vergleich zu SoSe 2022. Dies kann vorsichtig dahingehend interpretiert werden, dass die Aufgaben hoch korreliert sind und dass sie das gleiche zugrunde liegende Konzept messen. Im SoSe 2022 ist das Cronbachs Alpha für jede Aufgabe unterschiedlicher. Die Auswirkungen von Aufgabe 1 sind größer als bei allen anderen Aufgaben. Es ist jedoch anzumerken, dass diese Aufgabe doppelt so viele Punkte wie jede andere hat. Aufgabe 3 hat einen negativen Einfluss auf das Cronbachs Alpha und in Abbildung 7 zeigt diese Aufgabe einen niedrigen Diskriminationsindex. Es kann vermutet werden, dass die bei dieser Aufgabe erzielten Ergebnisse nicht mit dem Trend übereinstimmen, den anderen Aufgaben zeigten. Eine weitere Bewertung der Aufgabenbeschreibung wäre erforderlich, um festzustellen, ob die Aufgabe für Studierende unklar war.

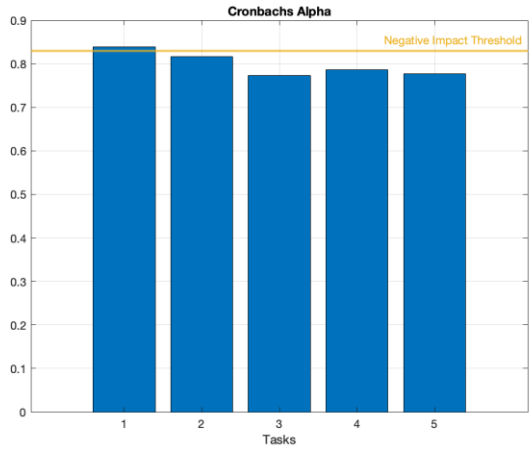


Abbildung 8. Cronbachs alpha SoSe 2021

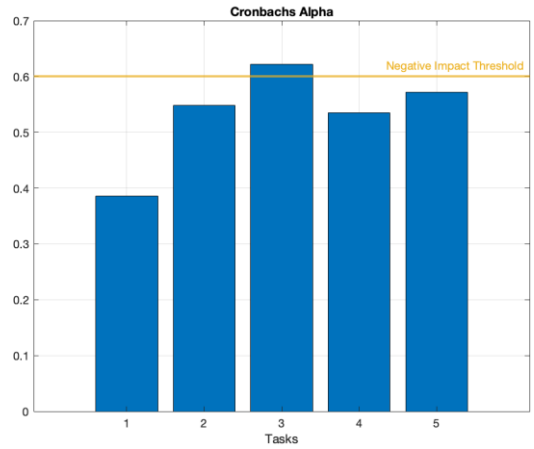


Abbildung 9. Cronbachs alpha SoSe 2022

Beschränkungen

Die Auswertung der Ergebnisse muss aufgrund der geringen Anzahl von Teilnehmenden vorsichtig erfolgen. In jedem Semester nahmen nur 32–33 Studierende an der Prüfung teil.

Zukünftige Analyse

Für die nächsten Prüfungen ist geplant, die Analyse unter Berücksichtigung einer feineren Unterscheidung der Prüfungsaufgaben durchzuführen. Auf diese Weise kann die Korrelation der Aufgaben bewertet werden, die auf ein ähnliches Niveau der Blooms Taxonomie abzielen. Außerdem ist für das nächste Semester geplant, Studierenden eine Version einer alten Prüfung zur Verfügung zu stellen, um sie bei der Vorbereitung zu unterstützen.

Script for the analysis of exam results

Loading the exam results

- The number of points collected should be discriminated by task and student.
- An ID for the students should NOT be included
- The first line is assumed to be the maximum number of points per task

```
 % csv file with points per task (5 tasks are assumed)
 clear

 fileName = 'SS22.csv';
 semester = fileName(1:end-4);
 fid = fopen(fileName,'r');
 count=0;
 pointsPerTask = [];
 while feof(fid) == 0
   count = count + 1;
   tmp=fscanf(fid,'%s;%s\n');
   if size(tmp)~=0
     pointsPerTask = cat(2, pointsPerTask, sscanf(strrep(tmp,',','.'),'
%f;'));
   end
 end
 fclose(fid);

 pointsPerTask=pointsPerTask.';
 maxPoints = pointsPerTask(1,:);
 numTasks = 5;
 pointsPerTask(1,:)=[];          % delete first row containing max points per
task
```

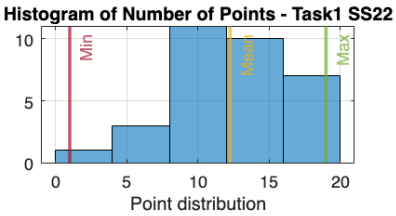
Metrics calculation

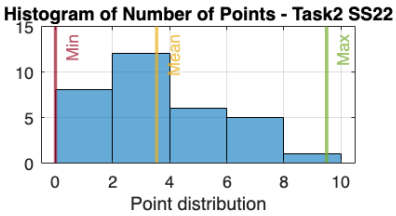
```
 % Maximum number of points per task
 maximum = max(pointsPerTask);

 % Minimum number of points per task
 minimum = min(pointsPerTask);

 % Average number of points per task
 avg = mean(pointsPerTask);

 % Task difficulty
 difficulty = avg./maxPoints;

 % Variance on the number of points per task
 variance = var(pointsPerTask);

 % Standard deviation of the number of points per task
 stdDeviation = std(pointsPerTask);

 % Discrimination power
 % total number of points per student
 sumPoints = sum(pointsPerTask,2);
 % sum number of points without each task
 partialSumPoints = sumPoints - pointsPerTask;
 discPower = diag(corr(pointsPerTask, partialSumPoints))';

 % Cronbach's Alpha
 cronbachsAlpha_all = (numTasks/(numTasks-1))*(1-
sum(variance)/var(sumPoints));
 cronbachsAlpha_task = ((numTasks-1)/(numTasks-2))*(1-(sum(variance)-
variance)./var(partialSumPoints));

 % Correlation
 correlationTasks = corr(pointsPerTask,pointsPerTask)
```

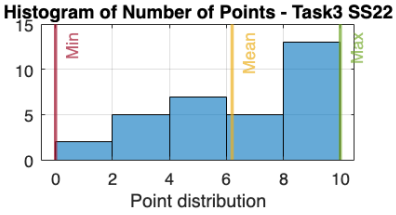
```
correlationTasks = 5×5

    1.0000    0.6022    0.2009    0.3855    0.2525
    0.6022    1.0000    0.1308    0.0838   -0.0803
    0.2009    0.1308    1.0000    0.0898    0.1839
    0.3855    0.0838    0.0898    1.0000    0.4590
    0.2525   -0.0803    0.1839    0.4590    1.0000
```

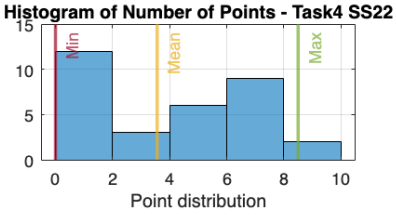
Plotting the results

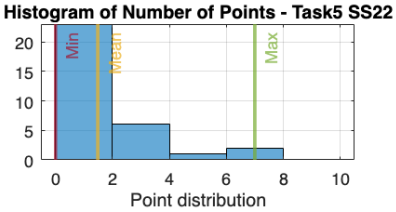
```
 % Histogram of number of points per task
 for t = 1:numTasks
     subplot(3,2,t)
     edges = 0:maxPoints(t)/5:maxPoints(t);
     histogram(pointsPerTask(:,t), edges)
     xline(minimum(t),'-','Min','Color','#A2142F','LineWidth',2)
     xline(avg(t),'-','Mean','Color','#EDB120','LineWidth',2)
     xline(maximum(t),'-','Max','Color','#77AC30','LineWidth',2)
     xlabel('Point distribution')
     title(strcat('Histogram of Number of Points - Task',num2str(t), {'
'},semester))
     grid on
 end


 % Histogram of number of points per task
 subplot(3,2,numTasks+1)
 edges = 0:sum(maxPoints)/5:sum(maxPoints);
 histogram(sum(pointsPerTask,2), edges)
 xline(min(sum(pointsPerTask,2)),'-','Min','Color','#A2142F','LineWidth',2)
 xline(mean(sum(pointsPerTask,2)),'-','Mean','Color','#EDB120','LineWidth',2)
 xline(max(sum(pointsPerTask,2)),'-','Max','Color','#77AC30','LineWidth',2)
 xlabel('Point distribution')
 title(strcat('Histogram of Total Number of Points', {' '}, semester))
 grid on
 for t = 1:numTasks
     subplot(3,2,t)
     edges = 0:maxPoints(t)/5:maxPoints(t);
     histogram(pointsPerTask(:,t), edges)
     xline(minimum(t),'-','Min','Color','#A2142F','LineWidth',2)
     xline(avg(t),'-','Mean','Color','#EDB120','LineWidth',2)
     xline(maximum(t),'-','Max','Color','#77AC30','LineWidth',2)
     xlabel('Point distribution')
     title(strcat('Histogram of Number of Points - Task',num2str(t), {' '},
semester))
     grid on
 end
```

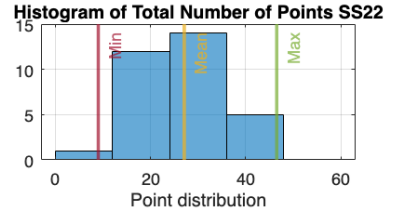

Histogram of Number of Points - Task1 SS22
Min
Mean
Max
Point distribution


Histogram of Number of Points - Task2 SS22
Min
Mean
Max
Point distribution

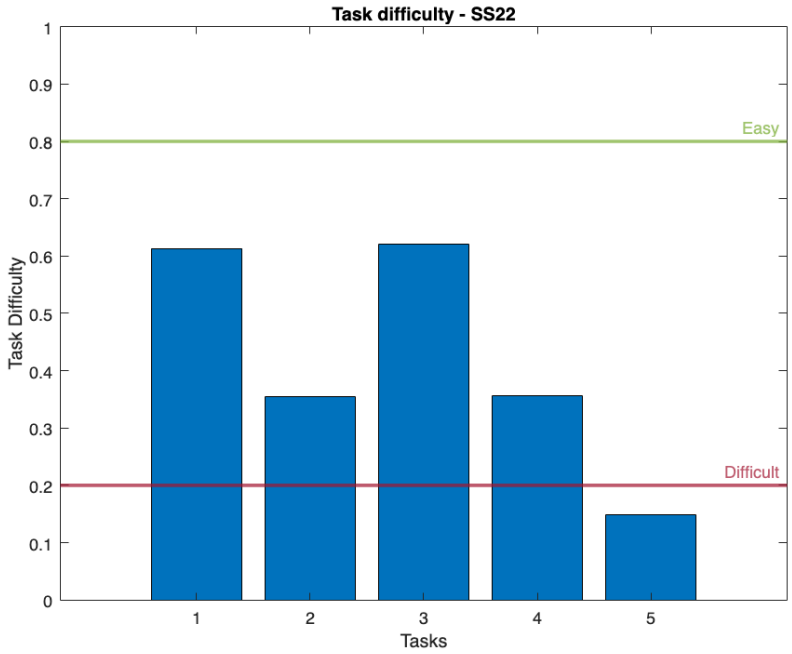

Histogram of Number of Points - Task3 SS22
Min
Mean
Max
Point distribution


Histogram of Number of Points - Task4 SS22
Min
Mean
Max
Point distribution

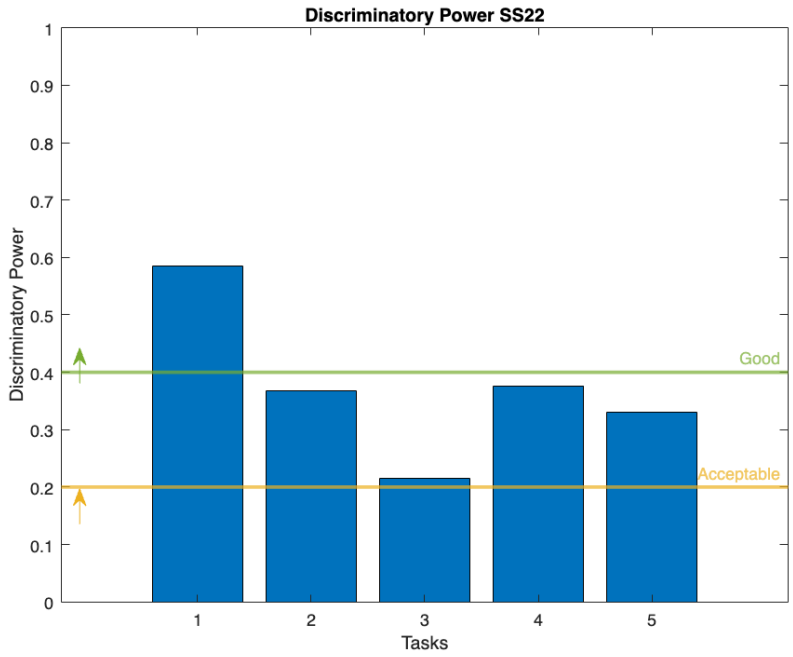

Histogram of Number of Points - Task5 SS22
Min
Mean
Max
Point distribution


Histogram of Total Number of Points SS22
Min
Mean
Max
Point distribution

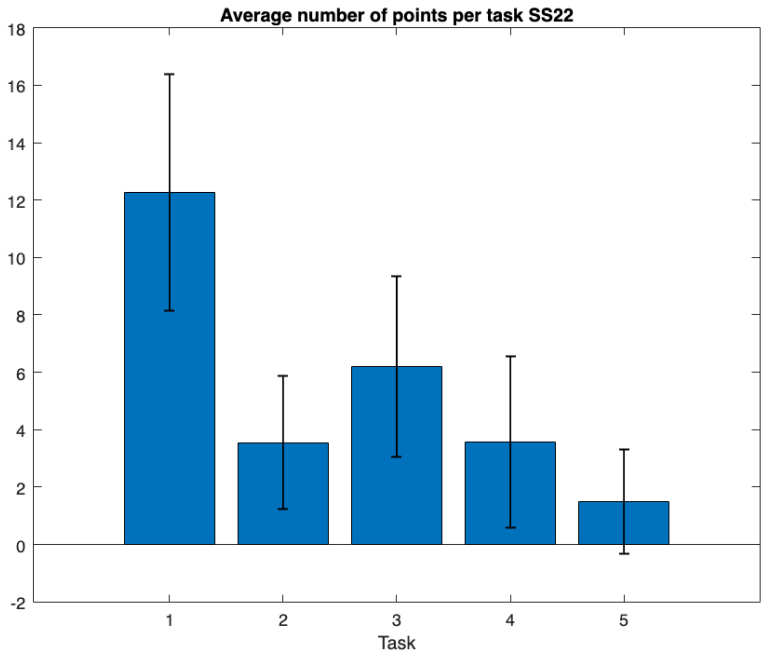
```
 % Difficulty
 figure, bar(difficulty)
 yline(0.8,'-','Easy','Color','#77AC30','LineWidth',2)
 yline(0.2,'-','Difficult','Color','#A2142F','LineWidth',2)
 ylim([0 1]), xlabel('Tasks'), ylabel('Task Difficulty'), title(strcat('Task
difficulty - ', {' '}, semester))
```



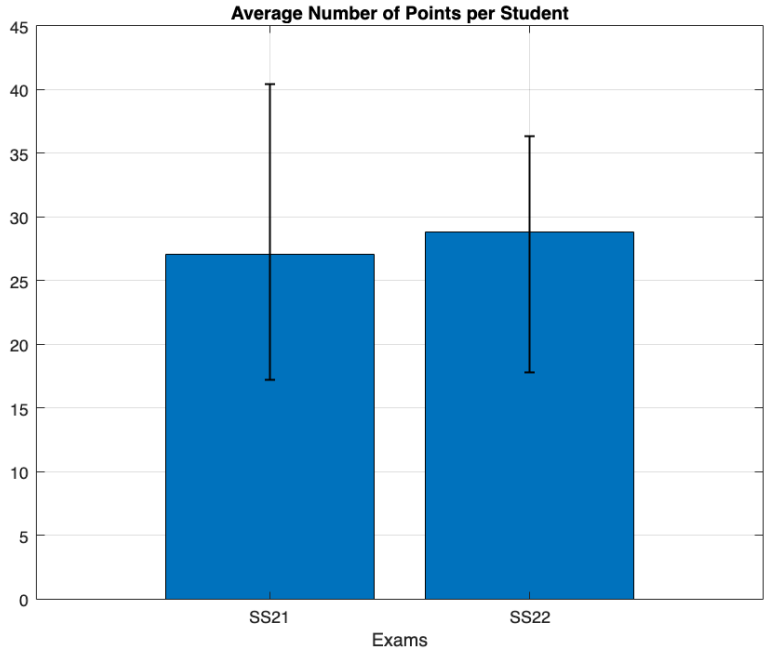
```
 % Discriminatory Power
 figure, bar(discPower)
 yline(0.4,'-','Good','Color','#77AC30','LineWidth',2)
 X = [0.15 0.15]; Y = [0.42  0.47];  annotation('arrow',X,Y,
'Color','#77AC30');
 yline(0.2,'-','Acceptable','Color','#EDB120','LineWidth',2)
 X = [0.15 0.15]; Y = [0.22  0.27];  annotation('arrow',X,Y,
'Color','#EDB120');
 ylim([0 1]), xlabel('Tasks'), ylabel('Discriminatory Power'),
title(strcat('Discriminatory Power', {' '}, semester))
```



```
% Mean-Variance Analysis
figure, bar(avg)
hold on
er = errorbar(avg,stdDeviation, LineWidth=1);
er.Color = [0 0 0];
er.LineStyle = 'none';  title(strcat('Average number of points per task', {'
'}, semester)); xlabel('Task')
```



```
 % Total Mean-Variance Analysis
 mean(sumPoints);
 std(sumPoints);
 comparisonSemester = 'SS21';
 valuesComparisonSemester = [28.84 11.60];        % first value is the mean,
second is the standard deviation
 X = categorical([{semester}, {comparisonSemester}]);
 figure, bar(X, [valuesComparisonSemester(1) mean(sumPoints)])
 hold on
 er = errorbar([valuesComparisonSemester(1)
mean(sumPoints)],[valuesComparisonSemester(2) std(sumPoints)], LineWidth=1);
 er.Color = [0 0 0];
 grid on
 er.LineStyle = 'none';  title('Average Number of Points per Student');
xlabel('Exams')
```



```
 % Cronbach's alpha
 figure, bar(cronbachsAlpha_task)
 yline(cronbachsAlpha_all,'-','Negative Impact
Threshold','Color','#EDB120','LineWidth',2)
 grid on; title(strcat('Cronbachs Alpha', {' '}, semester)); xlabel('Tasks');
```