AP® Statistics
2001 Sample Student Responses

The materials included in these files are intended for non-commercial use by AP teachers for course and exam preparation; permission for any other use must be sought from the Advanced Placement Program. Teachers may reproduce them, in whole or in part, in limited quantities, for face-to-face teaching purposes but may not mass distribute the materials, electronically or otherwise. These materials and any copies made of them may not be resold, and the copyright notices must be retained as they appear here. This permission does not apply to any third-party copyrights contained herein.

STATISTICS
SECTION II
Part B
Question 6
Spend about 25 minutes on this part of the exam.
Percent of Section II grade—25

Directions: Show all your work. Indicate clearly the methods you use, because you will be graded on the correctness of your methods as well as on the accuracy of your results and explanation.

6. The statistics department at a large university is trying to determine if it is possible to predict whether an applicant will successfully complete the Ph.D. program or will leave before completing the program. The department is considering whether GPA (grade point average) in undergraduate statistics and mathematics courses (a measure of performance) and mean number of credit hours per semester (a measure of workload) would be helpful measures. To gather data, a random sample of 20 entering students from the past 5 years is taken. The data are given below.

Successfully Completed Ph.D. Program

Student	A	B	C	D	E	F	G	H	I	J	K	L	M
GPA	3.8	3.5	4.0	3.9	2.9	3.5	3.5	4.0	3.9	3.0	3.4	3.7	3.6
Credit hours	12.7	13.1	12.5	13.0	15.0	14.7	14.5	12.0	13.1	15.3	14.6	12.5	14.0

Did Not Complete Ph.D. Program

Student	N	O	P	Q	R	S	T
GPA	3.6	2.9	3.1	3.5	3.9	3.6	3.3
Credit hours	11.1	14.5	14.0	10.9	11.5	12.1	12.0

The regression output at the top of the next page resulted from fitting a line to the data in each group. The residual plots (not shown) indicated no unusual patterns, and the assumptions necessary for inference were judged to be reasonable.

Successfully Completed Ph.D. Program

Predictor	Coef	StDev	T	P
Constant	23.514	1.684	13.95	0.000
GPA	−2.7555	0.4668	−5.90	0.000
S = 0.5658		R-Sq = 76.0%		

Did Not Complete Ph.D. Program

Predictor	Coef	StDev	T	P
Constant	24.200	3.474	6.97	0.001
GPA	−3.485	1.013	−3.44	0.018
S = 0.8408		R-Sq = 70.3%		

(a) Use an appropriate graphical display to compare the GPA's for the two groups. Write a few sentences commenting on your display.

GPA Comparison

Successful in   Not successful in
  Ph.D. program   PHD. Program

```
              2
         9  2  9
       4 0  3  1 3
 9 9 8 7 6 5 5 5  3  5 6 6 9
     0 0  4
          4
```
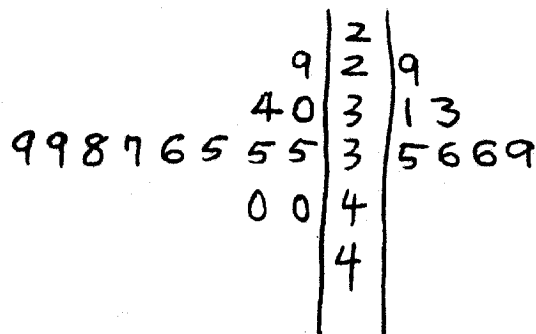
* the middle line numbers is the first digit of their GPA
Ex: 2.9   9|2|9

These who were successful in completing the Ph D program has a GPA distribution roughly symmetric and unimodel, and the range was from 2.9 - 4.0 (difference 1.1). The mean is somewhere around 3.7 and the median is 3.6.

Those who were not successful in completing the Ph D program has a GPA distribution skewed to the right (more in the upper 3. range). They range from 2.9 to 3.9 (1.0 difference). The mean is around 3.6 and the median was 3.5.

(b) For the students who successfully completed the Ph.D. program, is there a significant relationship between GPA and mean number of credit hours per semester?

Give a statistical justification to support your response.

Successful

$$\widehat{credit} = -2.7555 \,(GPA) + 23.514$$

$H_0: \beta = 0$ (no association)

$H_a: \beta \neq 0$
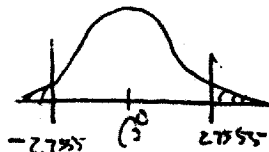
☑ Data shows linear relationship

☑ Risidual $\approx$ normal (not curved or patterned)
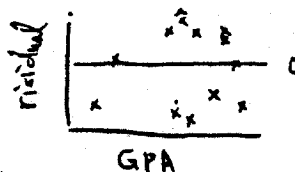
∴ use a β two-tailed t-test with d.f. $= 13-2 = 11$

$$P(\beta < -2.7555) = P\left(t < \frac{-2.7555-0}{0.4668}\right)$$

$$= P(t < -5.90)$$

$$\approx .00005$$

Reject H_0 since p-value $= .0001 < \alpha = .01$.

Therefore, there is a significant ~~relationship~~ relationship between GPA and the mean number of credit hours per semester.

(c) If a new applicant has a GPA of 3.5 and a mean number of credit hours per semester of 14.0, do you think this applicant will successfully complete the Ph.D. program? Give a statistical justification to support your response.

Successful

$\widehat{Credit} = -2.7555 (GPA) + 23.514$

$\widehat{credit} = -2.7555 (3.5) + 23.514$
$= 13.8698$

$\hat{y} - y = 13.87 - 14 = -.13$

Not successful

$\widehat{credit} = -3.485 (GPA) + 24.2$

$\widehat{credit} = -3.485 (3.5) + 24.2$
$= 12.0025$

$\hat{y} - y = 12 - 14 = -2$

I think this applicant will successfully complete the Ph. D. program because the successful regression better predicts his/her credit hours per semester than that the unsuccessful regression.

STATISTICS
SECTION II
Part B
Question 6
Spend about 25 minutes on this part of the exam.
Percent of Section II grade—25

Directions: Show all your work. Indicate clearly the methods you use, because you will be graded on the correctness of your methods as well as on the accuracy of your results and explanation.

6. The statistics department at a large university is trying to determine if it is possible to predict whether an applicant will successfully complete the Ph.D. program or will leave before completing the program. The department is considering whether GPA (grade point average) in undergraduate statistics and mathematics courses (a measure of performance) and mean number of credit hours per semester (a measure of workload) would be helpful measures. To gather data, a random sample of 20 entering students from the past 5 years is taken. The data are given below.

Successfully Completed Ph.D. Program

Student	A	B	C	D	E	F	G	H	I	J	K	L	M
GPA	3.8	3.5	4.0	3.9	2.9	3.5	3.5	4.0	3.9	3.0	3.4	3.7	3.6
Credit hours	12.7	13.1	12.5	13.0	15.0	14.7	14.5	12.0	13.1	15.3	14.6	12.5	14.0

Did Not Complete Ph.D. Program

Student	N	O	P	Q	R	S	T
GPA	3.6	2.9	3.1	3.5	3.9	3.6	3.3
Credit hours	11.1	14.5	14.0	10.9	11.5	12.1	12.0

The regression output at the top of the next page resulted from fitting a line to the data in each group. The residual plots (not shown) indicated no unusual patterns, and the assumptions necessary for inference were judged to be reasonable.

Successfully Completed Ph.D. Program

Predictor	Coef	StDev	T	P
Constant	23.514	1.684	13.95	0.000
GPA	−2.7555	0.4668	−5.90	0.000
S = 0.5658		R-Sq = 76.0%		

Did Not Complete Ph.D. Program

Predictor	Coef	StDev	T	P
Constant	24.200	3.474	6.97	0.001
GPA	−3.485	1.013	−3.44	0.018
S = 0.8408		R-Sq = 70.3%		

(a) Use an appropriate graphical display to compare the GPA's for the two groups. Write a few sentences commenting on your display.

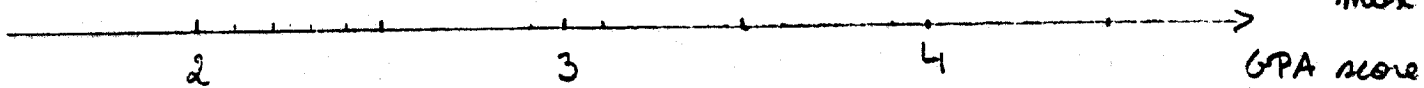5# summary completed
min 2.9
Q₁ : 3.45
Med : 3.6
Q₃ : 3.9
max : 4

5# summary did not
min : 2.9
Q₁ : 3.1
Med : 3.5
Q₃ : 3.6
max : 3.9
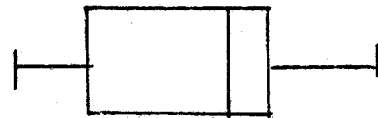
We can see that both groups have the same min. GPA score and approximately the same max GPA score.

We can see that students of the group that successfully completed the program tend to have higher GPA in general than those of the second group : their Q₁ is almost as high as the median of group 2.

→ In general we can say that students who completed the program have higher GPA score.

(b) For the students who successfully completed the Ph.D. program, is there a significant relationship between GPA and mean number of credit hours per semester?

Give a statistical justification to support your response.

Based on the regression output, the linear regression line has for equation:

$$(\text{Credit hours}) = 23.514 - 2.7555 \,(\text{GPA})$$

To be able to determine if the linear model is a good one to predict credit hours based on GPA we have to do a regression line t-test.

$H_0: \beta = 0$ β = true slope of the linear regression line

$H_A: \beta \neq 0$ $\alpha = 0.01$

$t = -5.9$

$P\text{-value} = 0.000$ $\Big\}$ given on regression output

Because the P-value is so small we reject H_0.
There is sufficient evidence to conclude that the LSRL is a good model in predicting credit hours based on GPA.

$\Rightarrow$ We can conclude that there is a significant linear relationship between GPA and mean number of credit hours per semester.
We can also add that the relation is negative.

(c) If a new applicant has a GPA of 3.5 and a mean number of credit hours per semester of 14.0, do you think this applicant will successfully complete the Ph.D. program? Give a statistical justification to support your response.

When I subtitute his GPA score in the equation of the LSRL of the students who completed the program I get an expected mean number of credit hours per semester of 13.87.

$$\hat{y} = 23.514 - 2.7555 \times 3.5$$
$$= 13.87$$

Because the predicted value is pretty close from the observed value (14), this new applicant seems to follow the pattern of the students who completed the program. Therefore I would say he will succeed.

STATISTICS
SECTION II
Part B
Question 6
Spend about 25 minutes on this part of the exam.
Percent of Section II grade—25

Directions: Show all your work. Indicate clearly the methods you use, because you will be graded on the correctness of your methods as well as on the accuracy of your results and explanation.

6. The statistics department at a large university is trying to determine if it is possible to predict whether an applicant will successfully complete the Ph.D. program or will leave before completing the program. The department is considering whether GPA (grade point average) in undergraduate statistics and mathematics courses (a measure of performance) and mean number of credit hours per semester (a measure of workload) would be helpful measures. To gather data, a random sample of 20 entering students from the past 5 years is taken. The data are given below.

Successfully Completed Ph.D. Program

Student	A	B	C	D	E	F	G	H	I	J	K	L	M
GPA	3.8	3.5	4.0	3.9	2.9	3.5	3.5	4.0	3.9	3.0	3.4	3.7	3.6
Credit hours	12.7	13.1	12.5	13.0	15.0	14.7	14.5	12.0	13.1	15.3	14.6	12.5	14.0

Did Not Complete Ph.D. Program

Student	N	O	P	Q	R	S	T
GPA	3.6	2.9	3.1	3.5	3.9	3.6	3.3
Credit hours	11.1	14.5	14.0	10.9	11.5	12.1	12.0

The regression output at the top of the next page resulted from fitting a line to the data in each group. The residual plots (not shown) indicated no unusual patterns, and the assumptions necessary for inference were judged to be reasonable.

Successfully Completed Ph.D. Program

Predictor	Coef	StDev	T	P
Constant	23.514	1.684	13.95	0.000
GPA	−2.7555	0.4668	−5.90	0.000
S = 0.5658		R-Sq = 76.0%		

Did Not Complete Ph.D. Program

Predictor	Coef	StDev	T	P
Constant	24.200	3.474	6.97	0.001
GPA	−3.485	1.013	−3.44	0.018
S = 0.8408		R-Sq = 70.3%		

(a) Use an appropriate graphical display to compare the GPA's for the two groups. Write a few sentences commenting on your display.

~~ŷ = credit hours~~
PhD completed

x = GPA Phd completed

~~constant = 23.514~~

~~ŷ = 23.514 − 2.7555X~~

~~ŷ = credit~~ hours, Phd not complete

x = GPA, Ph D not complete

~~constant = 24.200~~

~~ŷ = 24.200 − 3.485X~~
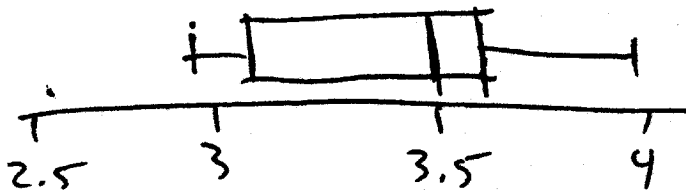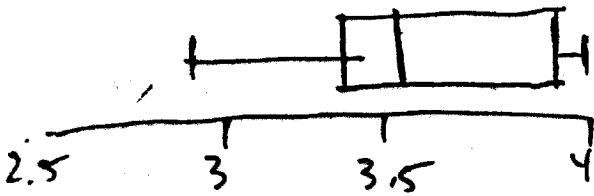
complete PhD   Not complete

Five number sumary

lo 2.9
Q₁ 3.45
med 3.6
Q₃ = 3.9
max = 4

lo 2.9
Q₁ = 3.1
med = 3.5
Q₃ = 3.6
max = 3.9

[box plot: complete PhD, axis 2.5, 3, 3.5, 4]

[box plot: Not complete, axis 2.5, 3, 3.5, 4]

As we can see, PhD complete has a much higher GPA median and quartiles, as well as max, than not complete PhD. Complete PhD is skewed to left, meaning it has mostly high GPA, while not complete is skewed to right, mostly lower GPA.

(b) For the students who successfully completed the Ph.D. program, is there a significant relationship between GPA and mean number of credit hours per semester?

Give a statistical justification to support your response.

For the students who successfully completed the PhD program, there is a significant relationship between GPA and credit hours. R-squared value is .76, and $r = \sqrt{r^2} = \sqrt{.76} = .87177$. The results, though, give a negative value for GPA (-2.7555). Therefore, the relationship is fairly close linear, however, it is a negative association because as GPA increases, the number of credit hours decreases. (slope of least squares reg line = -2.7555).

(c) If a new applicant has a GPA of 3.5 and a mean number of credit hours per semester of 14.0, do you think this applicant will successfully complete the Ph.D. program? Give a statistical justification to support your response.

We will get the least squares regression line for both comple and not complete phD program, and see which ~~it is less to predict if he~~ will complete it. (predicted credit hours from GPA=3.5 is closer to 14.)

PhD

slope $= -2.7555$
$x = GPA$
constant $= 23.514$
$\hat{y} =$ credit hours

$\hat{y} = 23.514 - 2.7555 x$

$\hat{y} = 23.514 - 2.7555 (3.5)$

$\hat{y} = 13.86975$

No PhD

slope $= -3.485$
$x = GPA$
constant $= 24.200$
$\hat{y} =$ credit hours

$\hat{y} = 24.200 - 3.485 x$

$\hat{y} = 24.200 - 3.485 (3.5)$

$\hat{y} = 12.0025$

The predicted number of credit hours (actual is 14), from a GPA of 3.5, is much closer to 14 for those who complete PhD program than for those who do not. This suggests that this student is likely to be successful in the PhD program.