Stat 13, Tue 6/5/12.
1. Hand in hw7.
2. Practice problems.

Final exam is Thur, 6/7, in class.
It has 20 multiple choice questions. Only about 50% are on regression and correlation. The other 50% are on previous topics like study design and confounding factors (10%), probability (10%), normal calculations (10%), standard errors and confidence intervals (10%), and testing (10%). All these percentages here are approximate.

You will get no credit for saying “none of the above” unless your answer is exactly right.
You can use 2 pages, each double sided, of notes. Bring a calculator and a pen or pencil. Do not hand in your notes when you hand in the exam.

1. Suppose you are trying to use regression to predict wine prices by taste ratings. You take a SRS of 300 wines, have people rate their taste from 0 to 100, and record the prices of the wines. Suppose both taste ratings and wine prices are normally distributed. Suppose you find that your mean wine price is \$10 and the sd is \$4.

a. What percentage of wines cost more than \$15?

Standardize the question: (\$15 - \$10)/\$4 = 1.25, so we want the area under the standard normal curve greater than 1.25.
Using the z table, this is 1 - 0.8944 = 0.1056 = 10.56%.

1. Suppose you are trying to use regression to predict wine prices by taste ratings. You take a SRS of 300 wines, have people rate their taste from 0 to 100, and record the prices of the wines. Suppose both taste ratings and wine prices are normally distributed. Suppose you find that your mean wine price is $10 and the sd is $4.

b. 95% of wine prices are in what range?

$10 +/- 1.96 ($4) = $10 +/- $7.84.

1. Suppose you are trying to use regression to predict wine prices by taste ratings. You take a SRS of 300 wines, have people rate their taste from 0 to 100, and record the prices of the wines. Suppose both taste ratings and wine prices are normally distributed. Suppose you find that your mean wine price is $10 and the sd is $4.

c. What price would put a wine in the 80th percentile of price?

Using the z table, the 80th percentile of the standard normal is about 0.84. We need to convert this 0.84 from standard units to dollars.

0.84 ($4) + $10 = $13.36.

1. Suppose you are trying to use regression to predict wine prices by taste ratings. You take a SRS of 300 wines, have people rate their taste from 0 to 100, and record the prices of the wines. Suppose both taste ratings and wine prices are normally distributed. Suppose you find that your mean wine price is $10 and the sd is $4.

d. Suppose the slope of the regression line is .1, and the correlation is 0.5. What is the sd of the taste ratings in your sample?

$b = r\ s_y / s_x$,
so $s_x = r\ s_y / b = (0.5)(4)/.1 \sim 20$.

1. Suppose you are trying to use regression to predict wine prices by taste ratings. You take a SRS of 300 wines, have people rate their taste from 0 to 100, and record the prices of the wines. Suppose both taste ratings and wine prices are normally distributed. Suppose you find that your mean wine price is $10 and the sd is $4.

Suppose the slope of the regression line is .1, and the correlation is 0.5.

e. Suppose the intercept of the regression line is $6. What is the sample mean wine taste rating?

b = .1, and we know a = $\bar{y}$ - b $\bar{x}$ = 10 - .1 $\bar{x}$ = $6.
So, .1 $\bar{x}$ = 10 - 6 = 4, so $\bar{x}$ = 4/.1 = 40.

1. Suppose you are trying to use regression to predict wine prices by taste ratings. You take a SRS of 300 wines, have people rate their taste from 0 to 100, and record the prices of the wines. Suppose both taste ratings and wine prices are normally distributed. Suppose you find that your mean wine price is $10 and the sd is $4.

Suppose again that the slope of the regression line is .1, the correlation is 0.5, and the intercept of the regression line is $6.

f. Use the regression line to predict the cost of a wine that has a taste rating of 55.

b = .1, and a = $6, so $\hat{y}$= $6 + .1 x, and here x = 55, so
$\hat{y}$ = $6 + .1 (55) = $11.50.

1. Suppose you are trying to use regression to predict wine prices by taste ratings. You take a SRS of 300 wines, have people rate their taste from 0 to 100, and record the prices of the wines. Suppose both taste ratings and wine prices are normally distributed. Suppose you find that your mean wine price is \$10 and the sd is \$4.

Suppose again that the slope of the regression line is .1, the correlation is 0.5, and the intercept of the regression line is \$6.

g. +/- what? That is, roughly how much do you expect your prediction to be off by?

$\sqrt{(1-r^2)}\, s_y = \sqrt{1-.5^2}$ (\$4) ~ \$3.46.

1. Suppose you are trying to use regression to predict wine prices by taste ratings. You take a SRS of 300 wines, have people rate their taste from 0 to 100, and record the prices of the wines. Suppose both taste ratings and wine prices are normally distributed. Suppose you find that your mean wine price is $10 and the sd is $4.

Suppose again that the slope of the regression line is .1 and the intercept of the regression line is $6.

h. Which of the following is your interpretation of the regression estimates?

(i) If you make a wine taste 1 unit better, then it will cost ten cents more.
(ii) Wines that taste 1 unit better cost, on average, 10 cents more.
(iii) A wine that gets a taste score of 200 would cost around $6 + .1(200) = $26.

2. Suppose we deal two cards without replacement from an ordinary deck, and consider a face card a J, Q or K. What is the expected number of face cards you are dealt?

Let X = the number of face cards you are dealt. X = 0, 1, or 2.

E(X) = 0 P(X=0) + 1 P(X=1) + 2 P(X=2)

= 0 + 1P(X=1) + 2P(X=2).

P(X=1) = P(you get 1 face card and 1 nonface card)

= P(your 1st card is face and 2^{nd} isn't) + P(your 1^{st} card isn't face and your 2^{nd} is)

= P(1^{st} is face) P(2^{nd} isn't face | 1^{st} is face) + P(1^{st} isn't face)P(2^{nd} is face | 1^{st} isn't)

= 12/52 x 40/51 + 40/52 x 12/51

= 36.2%.

P(X=2) = P(1^{st} card is face and 2^{nd} is face)

= P(1^{st} card is face) P(2^{nd} is face | 1^{st} is face)

= 12/52 x 11/51

= 4.98%.

So E(X) = 0 + 1 x 36.2% + 2 x 4.98% = 0.4616.