ST512

Summer Session II, 2008

Quiz 1

Name:____

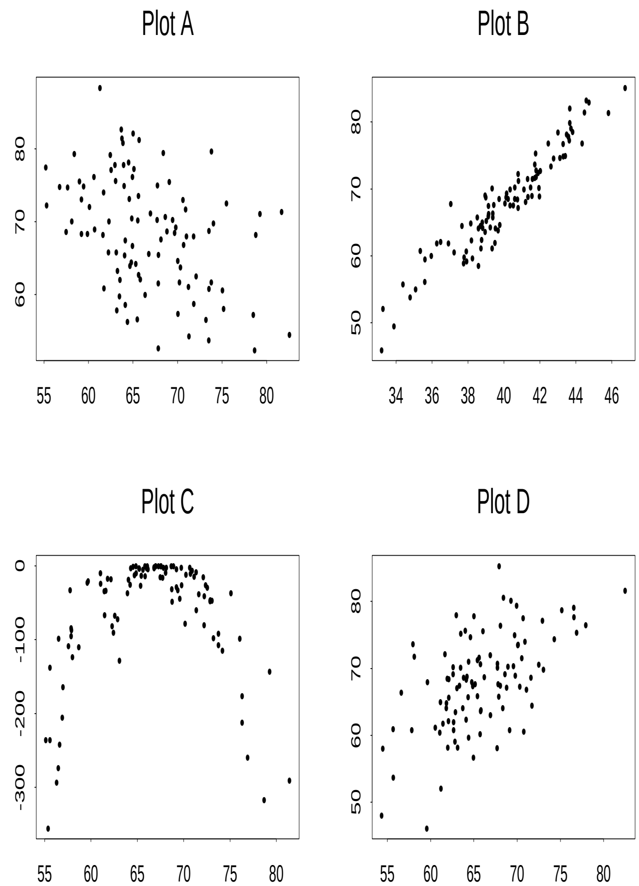
Directions: Answer questions as directed. Please show work. Give expressions for answers where possible, as partial credit may be awarded in cases where expressions are correct, but numerical answers are not.

You may use the back of the page if you need extra space.

1. (20 points) Four scatterplots appear below.

   (a) For each sample correlation coefficient, circle the letter of the corresponding plot (one letter per correlation coefficient).

   - $r = 0.58$ (A B C D)
   - $r = 0.95$ (A B C D)
   - $r = -0.41$ (A B C D)
   - $r = 0.13$ (A B C D)



   (b) For plot B, what proportion of variation on the vertical (y) axis would be explained by a linear regression on the variable on the horizontal (x) axis (a linear regression of y on x)?

2. (40 points) The crop yield of grapes (y, in tons/acre) harvested in August can be predicted using counts of berry clusters (x), measured in July. Data and output (from `PROC REG`) for a simple linear regression analysis are given at the end of the problem.

   (a) Give the simple linear model being considered here, establishing appropriate notation and specifying distributional assumptions.

   (b) Is there evidence that cluster count makes for a useful predictor? Find in the output a p-value for a test of the hypothesis that the mean yield does not depend on cluster count. Draw a brief conclusion.

(c) As the number of clusters increases by one unit, how much does the yield increase, on average? Report a 95% confidence interval for this quantity.

(d) As the number of clusters increases by 10 units, how much does the yield increase, on average? Report an estimate.

(e) The residual for the observation in 1975 is hidden. Deduce its value.

(f) A hurricane wiped out the 1972 crop. For insurance purposes, an estimate of what the yield would have been is needed. Given that the cluster count was $x = 125.0$, provide a 95% prediction interval for what the yield would have been.

year	cluscount	yield
1971	116.37	5.6
1973	82.77	3.2
1974	110.68	4.5
1975	97.50	4.2
1976	115.88	5.2
1978	125.24	4.8
1979	116.15	4.9

The REG Procedure

Descriptive Statistics

Variable	Sum	Mean	Uncorrected SS	Variance	Standard Deviation
Intercept	7.00000	1.00000	7.00000	0	0
cluscount	764.59000	109.22714	84753	206.53926	14.37147
yield	32.40000	4.62857	153.58000	0.60238	0.77613

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	2.71454	2.71454	15.08	0.0116
Error	5	0.89975	0.17995		
Corrected Total	6	3.61429			

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	-0.48355	1.32595	-0.36	0.7303
cluscount	1	0.04680	0.01205	3.88	0.0116

Output Statistics

Obs	Dependent Variable	Predicted Value	Residual
1	5.6000	4.9629	0.6371
2	3.2000	3.3903	-0.1903
3	4.5000	4.6966	-0.1966
4	4.2000	4.0797	xxxxxxx
5	5.2000	4.9399	0.2601
6	4.8000	5.3780	-0.5780
7	4.9000	4.9526	-0.0526

3. (40 points) In a random sample of $n = 40$ children, a simple linear regression of weight in kg (y) on height in cm (x) gives an error mean square of $MS[E] = 51.7$. The least squares regression line and sample means and standard deviations are given below:

$$\text{LS Regression line:} \quad \hat{y} = -57.67 + 0.66x.$$

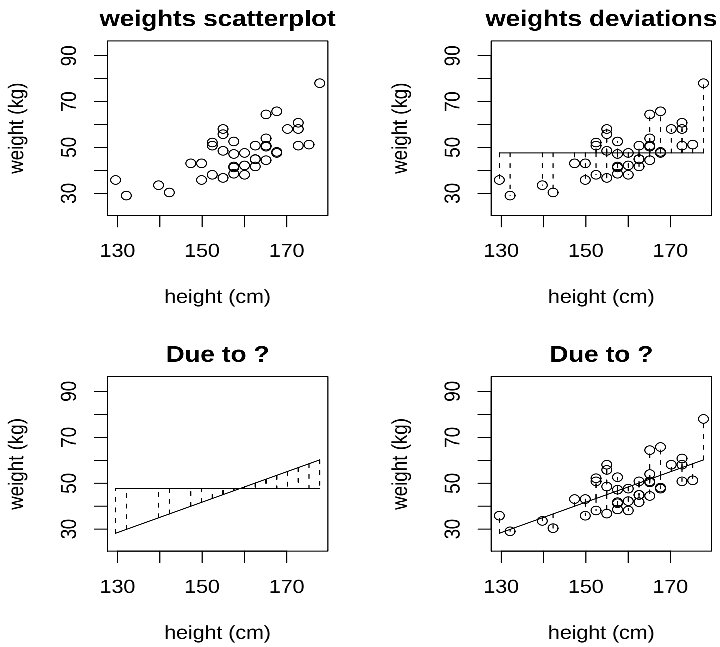
$$\begin{aligned} \bar{x} &= 158.90, & s_x &= 10.78 \\ \bar{y} &= 47.63, & s_y &= 10.07. \end{aligned}$$

(a) Using the linear regression, estimate the mean weight among all kids who have a height of $x = 158.9$ cm along with a standard error.

(b) Report the standard error associated with the slope estimate, $\hat{\beta}_1 = 0.66$.

(c) Deduce the value of the sample correlation coefficient, r.

(d) In the matrix of graphs below, the titles for the two plots at the bottom are incomplete. Complete each with a single word that describes the vertical distances depicted in these bottom two plots.



(e) Consider the population of shorter kids with a height of $x = 150cm$. Estimate the standard deviation of weights in this population.

Table of critical values ($t(df, \alpha)$) from t-distributions:

df	$\alpha = 0.2$	$\alpha = 0.15$	$\alpha = 0.1$	$\alpha = 0.05$	$\alpha = 0.025$	$\alpha = 0.01$	$\alpha = 0.005$	$\alpha = 0.001$
1	1.37638	1.96261	3.07768	6.31375	12.7062	31.8205	63.6567	318.309
2	1.06066	1.38621	1.88562	2.91999	4.3027	6.9646	9.9248	22.327
3	0.97847	1.24978	1.63774	2.35336	3.1824	4.5407	5.8409	10.215
4	0.94096	1.18957	1.53321	2.13185	2.7764	3.7469	4.6041	7.173
5	0.91954	1.15577	1.47588	2.01505	2.5706	3.3649	4.0321	5.893
6	0.90570	1.13416	1.43976	1.94318	2.4469	3.1427	3.7074	5.208
7	0.89603	1.11916	1.41492	1.89458	2.3646	2.9980	3.4995	4.785
8	0.88889	1.10815	1.39682	1.85955	2.3060	2.8965	3.3554	4.501
9	0.88340	1.09972	1.38303	1.83311	2.2622	2.8214	3.2498	4.297
10	0.87906	1.09306	1.37218	1.81246	2.2281	2.7638	3.1693	4.144
11	0.87553	1.08767	1.36343	1.79588	2.2010	2.7181	3.1058	4.025
12	0.87261	1.08321	1.35622	1.78229	2.1788	2.6810	3.0545	3.930
13	0.87015	1.07947	1.35017	1.77093	2.1604	2.6503	3.0123	3.852
14	0.86805	1.07628	1.34503	1.76131	2.1448	2.6245	2.9768	3.787
15	0.86624	1.07353	1.34061	1.75305	2.1314	2.6025	2.9467	3.733
16	0.86467	1.07114	1.33676	1.74588	2.1199	2.5835	2.9208	3.686
17	0.86328	1.06903	1.33338	1.73961	2.1098	2.5669	2.8982	3.646
18	0.86205	1.06717	1.33039	1.73406	2.1009	2.5524	2.8784	3.610
19	0.86095	1.06551	1.32773	1.72913	2.0930	2.5395	2.8609	3.579
20	0.85996	1.06402	1.32534	1.72472	2.0860	2.5280	2.8453	3.552
21	0.85907	1.06267	1.32319	1.72074	2.0796	2.5176	2.8314	3.527
22	0.85827	1.06145	1.32124	1.71714	2.0739	2.5083	2.8188	3.505
23	0.85753	1.06034	1.31946	1.71387	2.0687	2.4999	2.8073	3.485
24	0.85686	1.05932	1.31784	1.71088	2.0639	2.4922	2.7969	3.467
25	0.85624	1.05838	1.31635	1.70814	2.0595	2.4851	2.7874	3.450
26	0.85567	1.05752	1.31497	1.70562	2.0555	2.4786	2.7787	3.435
27	0.85514	1.05673	1.31370	1.70329	2.0518	2.4727	2.7707	3.421
28	0.85465	1.05599	1.31253	1.70113	2.0484	2.4671	2.7633	3.408
29	0.85419	1.05530	1.31143	1.69913	2.0452	2.4620	2.7564	3.396
30	0.85377	1.05466	1.31042	1.69726	2.0423	2.4573	2.7500	3.385
40	0.85070	1.05005	1.30308	1.68385	2.0211	2.4233	2.7045	3.307
60	0.84765	1.04547	1.29582	1.67065	2.0003	2.3901	2.6603	3.232
80	0.84614	1.04320	1.29222	1.66412	1.9901	2.3739	2.6387	3.195
100	0.84523	1.04184	1.29007	1.66023	1.9840	2.3642	2.6259	3.174
200	0.84342	1.03913	1.28580	1.65251	1.9719	2.3451	2.6006	3.131
100000	0.84162	1.03644	1.28156	1.64487	1.9600	2.3264	2.5759	3.090