MATH 50 FALL 2017

PRACTICE SESSION

- Note: For coefficients, we refer to least square estimation unless otherwise specified.

(1) In simple linear regression does the slope of the fitted line vary from one sample set to another?

(2) What can you say about this deviation?

(3) Given a linear model $z = \beta_0 + \beta_1 x + \varepsilon$, with assumptions as before. Suppose that you know exactly what are β_0,β_1 and σ.

 (a) Can you exactly calculate z for a given x ?

 (b) Can you give a prediction interval for z without doing regression analysis ?

(4) In simple regression we assumed that $\varepsilon_i \sim NID(o, \sigma^2)$. Suppose this is not true and $Var(\varepsilon_i)$ is not constant such that it decreases as x increases. Plot scatter diagram of a sample (of 10-20 points) that demonstrates such a behaviour. (For simplicity you can suppose your fitted coefficients are $\beta_0 = 0$, $\beta_1 = 1$).

(5) [T/F] $\hat{\sigma}^2$ is a n unbiased estimator of σ^2

(6) Is it possible that $\hat{\sigma}^2$ is significantly less than σ^2?

(7) (*) At what x is the CI for E(y|x) narrowest? At that particular point, what happens to the length of that interval as sample size n tends to infinity? How do you explain this?

(8) (*) At what x is the PI for y narrowest? At that particular point, what happens to the length of that interval as sample size n tends to infinity? How do you explain this? (note: as $dof \to \infty$ t-distribution approaches to the normal distribution).

(9) In simple linear regression (for $y \sim x$) the fitted regression line L is of the form ___________________, and the line L can be used

 [T/F]: to predict the value of y at a given x

 [T/F]: to visually analyze and check if relation is linear

 [T/F]: to estimate the rate of change in y as x decreases

(10) In simple linear regression, suppose we replaced observed values y_i with $y_i - C_0$ for some constant C_0. What happens to the coefficients ?

(11) How about in multiple linear regression ?

(12) In simple linear regression, suppose we replaced x_i with $x_i + C_0$ for some constant C_0. What happens to the fitted line ?

(13) How about in multiple linear regression ?

(14) Using no intercept model, can you obtain a simple regression model that passes through a given point (x_0, y_0)? (Hint consider no-intercept model as the model that passes through $(0, 0)$).

For the next 6-question consider multiple regression model

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \varepsilon_i,$$

where $\varepsilon_i \sim NID(o, \sigma^2)$. From 20 observations above model is fitted. Following results obtained.

	Coeff. Estimate	$se\left(\hat{\beta}_i\right)$
$\hat{\beta}_0$	6.1	5.0
$\hat{\beta}_1$	-2.2	1.1
$\hat{\beta}_2$	31.2	6
$\hat{\beta}_3$	5.1	2.0

$SS_T \approx 101.6 \quad SS_R \approx 77.1$

Answer below questions, otherwise if the given information is not sufficient, then state that.

(15) What is degrees of freedom ?

(16) Compute test statistic F_0 for significance of regression

(17) Test for significance of individual regression coefficients ($\alpha = 0.05$). Give an interpretation for these results.

(18) What is the Prob($\beta_3 > 2.58$) ?

(19) Estimate σ^2

(20) What is the smallest value G_1 such that the statement

$$\beta_3 > G_1 \qquad with\ confidence\ level\ 99\%$$

is rejected.

(21) .

1. $\hat{y} = 100 + 0.2x_1 + 4x_2$
2. $\hat{y} = 95 + 0.15x_1 + 3x_2 + 1x_1x_2$

Both]models have been built over the range $20 \leq x_1 \leq 50$ (°C) and $0.5 \leq x_2 \leq 10$ (hours).

a. Using both models, what is the predicted value of conversion when $x_2 = 2$ in terms of x_1? Repeat this calculation for $x_2 = 8$. Draw a graph of the predicted values as a function of temperature for both conversion models. Comment on the effect of the interaction term in model 2.

b. Find the expected change in the mean conversion for a unit change in temperature x_1 for model 1 when $x_2 = 5$. Does this quantity depend on the specific value of reaction time selected? Why?

c. Find the expected change in the mean conversion for a unit change in temperature x_1 for model 2 when $x_2 = 5$. Repeat this calculation for $x_2 = 2$ and $x_2 = 8$. Does the result depend on the value selected for x_2? Why?

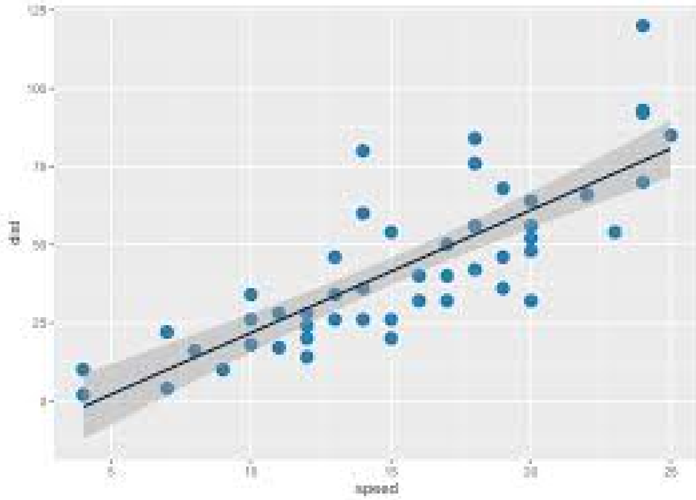
(22) In extra sum of squares method, we estimated contribution of a subset of regressors using sum of squares due to those regressors. Is it possible that $SS_R(\vec{\beta}_2|\vec{\beta}_1)$ is very small (and test statistic $F_0$is also small) but the regressors corresponding to coefficients in $\vec{\beta}_2$ has a strong cause-effect relationship with the response variable y.

(23) Let $\vec{\beta} = \begin{bmatrix} \vec{\beta}_1 \\ \vec{\beta}_2 \end{bmatrix}$. Is it true that $SS_R(\vec{\beta}_1|\vec{\beta}_2) + SS_R(\vec{\beta}_2|\vec{\beta}_1) = SS_R(\vec{\beta})$. Under what condition this might be true?

(24) For $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \varepsilon$. Using a sample data of size 3, give an example of hidden extrapolation. (also give a sketch of $x_1 - x_2$ plane)

(25) Suppose $Covar(\beta_0, \beta_1) = 0$, draw few possible shapes for 95% joint confidence region of $\vec{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix}$.

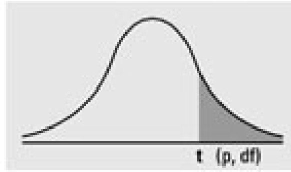
(26) Is it likely that below plot be 90% prediction interval of y? Why?



.

(27) From book (2.24, 2.29, 3.30, 3.21)

Numbers in each row of the table are values on a *t*-distribution with (*df*) degrees of freedom for selected right-tail (greater-than) probabilities (*p*).



df/p	0.40	0.25	0.10	0.05	0.025	0.01	0.005	0.0005
1	0.324920	1.000000	3.077684	6.313752	12.70620	31.82052	63.65674	636.6192
2	0.288675	0.816497	1.885618	2.919986	4.30265	6.96456	9.92484	31.5991
3	0.276671	0.764892	1.637744	2.353363	3.18245	4.54070	5.84091	12.9240
4	0.270722	0.740697	1.533206	2.131847	2.77645	3.74695	4.60409	8.6103
5	0.267181	0.726687	1.475884	2.015048	2.57058	3.36493	4.03214	6.8688
6	0.264835	0.717558	1.439756	1.943180	2.44691	3.14267	3.70743	5.9588
7	0.263167	0.711142	1.414924	1.894579	2.36462	2.99795	3.49948	5.4079
8	0.261921	0.706387	1.396815	1.859548	2.30600	2.89646	3.35539	5.0413
9	0.260955	0.702722	1.383029	1.833113	2.26216	2.82144	3.24984	4.7809
10	0.260185	0.699812	1.372184	1.812461	2.22814	2.76377	3.16927	4.5869
11	0.259556	0.697445	1.363430	1.795885	2.20099	2.71808	3.10581	4.4370
12	0.259033	0.695483	1.356217	1.782288	2.17881	2.68100	3.05454	43178
13	0.258591	0.693829	1.350171	1.770933	2.16037	2.65031	3.01228	4.2208
14	0.258213	0.692417	1.345030	1.761310	2.14479	2.62449	2.97684	4.1405
15	0.257885	0.691197	1.340606	1.753050	2.13145	2.60248	2.94671	4.0728
16	0.257599	0.690132	1.336757	1.745884	2.11991	2.58349	2.92078	4.0150
17	0.257347	0.689195	1.333379	1.739607	2.10982	2.56693	2.89823	3.9651
18	0.257123	0.688364	1.330391	1.734064	2.10092	2.55238	2.87844	3.9216
19	0.256923	0.687621	1.327728	1.729133	2.09302	2.53948	2.86093	3.8834
20	0.256743	0.686954	1.325341	1.724718	2.08596	2.52798	2.84534	3.8495
21	0.256580	0.686352	1.323188	1.720743	2.07961	2.51765	2.83136	3.8193
22	0.256432	0.685805	1.321237	1.717144	2.07387	2.50832	2.81876	3.7921
23	0.256297	0.685306	1.319460	1.713872	2.06866	2.49987	2.80734	3.7676
24	0.256173	0.684850	1.317836	1.710882	2.06390	2.49216	2.79694	3.7454
25	0.256060	0.684430	1.316345	1.708141	2.05954	2.48511	2.78744	3.7251
26	0.255955	0.684043	1.314972	1.705618	2.05553	2.47863	2.77871	3.7066
27	0.255858	0.683685	1.313703	1.703288	2.05183	2.47266	2.77068	3.6896
28	0.255768	0.683353	1.312527	1.701131	2.04841	2.46714	2.76326	3.6739
29	0.255684	0.683044	1.311434	1.699127	2.04523	2.46202	2.75639	3.6594
30	0.255605	0.682756	1.310415	1.697261	2.04227	2.45726	2.75000	3.6460
z	0.253347	0.674490	1.281552	1.644854	1.95996	2.32635	2.57583	3.2905
CI	———	———	80%	90%	95%	98%	99%	99.9%