The first problems are taken from previous final exams, with solutions at the end of this document. The other problems are suggested practice problems from the textbook, with solution keys at the end of the textbook.

1. A scientist wishes to compare two methods for removing atrazine contamination from the soil. Suppose he has available 8 contaminated plots of soil, each a square 10 meters on a side. Of the 8 plots, 4 are randomly assigned to receive method A and 4 to receive method B. On each plot, 3 (independent) readings are obtained. The following are the observed concentrations of atrazine in parts per million for each reading on each plot.

Method A				Method B			
Plot 1	Plot 2	Plot 3	Plot 4	Plot 5	Plot 6	Plot 7	Plot 8
4.9	6.5	5.5	5.8	6.5	7.2	5.6	5.9
5.2	6.3	5.0	5.7	6.0	7.0	5.4	6.3
4.9	6.4	5.4	5.6	6.4	7.1	5.5	6.1

   The scientist plans on analyzing the data with an independent sample t-test, by pooling the 12 observations on each method together as one sample for each method. Is this analysis correct? Justify briefly.

2. Thrombolytic drugs are substances that cause the breakdown of blood clots obstructing the flow of blood through the vessels. They are used by injection during or shortly after a heart attack or stroke to prevent clots from blocking blood flow to the heart muscle or brain. In an experiment, three thrombolitic treatments were studied: streptokinase (SK), accelerated alteplase (AtPA) and streptokinase plus alteplase (SK+tPA). A total of 4082 patients affected by Acute Myocardial Infarction (AMI) participated in the study. Each patient received one of the drugs, randomly assigned, shortly after the infarction. Efficacy was measured by the mortality at 30 to 35 days.

   (a) The data are summarized in the following table. Test the null hypothesis that the three thrombolitic drugs yield the same probability of survival (use $\alpha = 0.05$).

Drug	SK	AtPA	SK+tPA
Death	147	65	72
Survival	1869	969	960
Total	2016	1034	1032

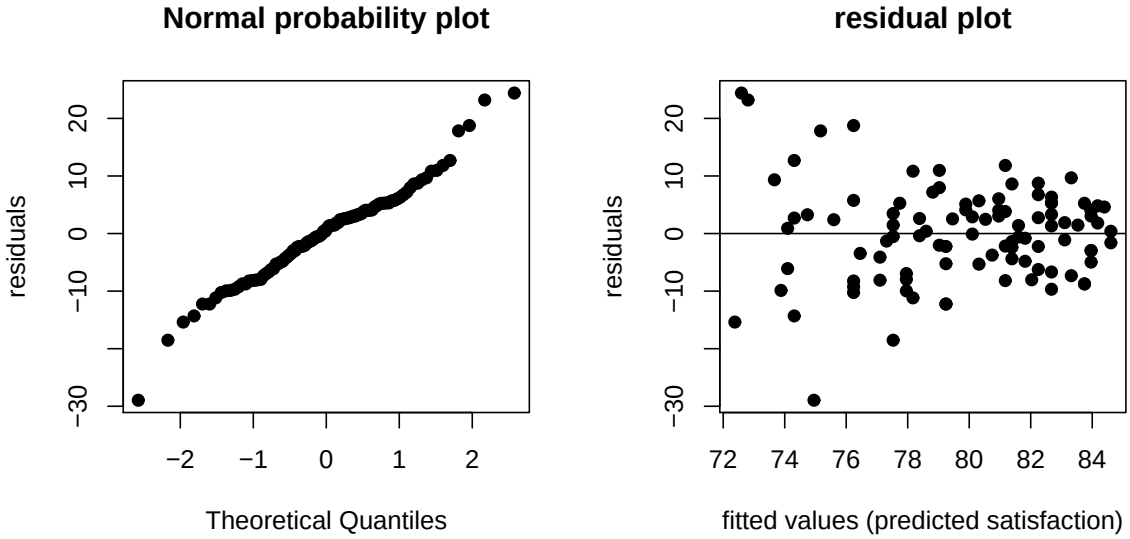
   (b) Check the validity of the method applied in 2a.

3. An investigation is performed to understand the relationship between age and satisfaction with medical care. Each individual in the investigation is asked to rate satisfaction with medical care on a scale of 0 to 100, with 100 denoting complete satisfaction. Age was also recorded (among other sociodemographic characteristics) on each individual. A total of 100 individuals are involved in the investigation. The linear regression analysis of satisfaction versus age yields satisfaction = 67.6 + 0.21 age, with a correlation coefficient $r = 0.37$ and slope $b_1 = 0.21$ (p-value = 0.0002).

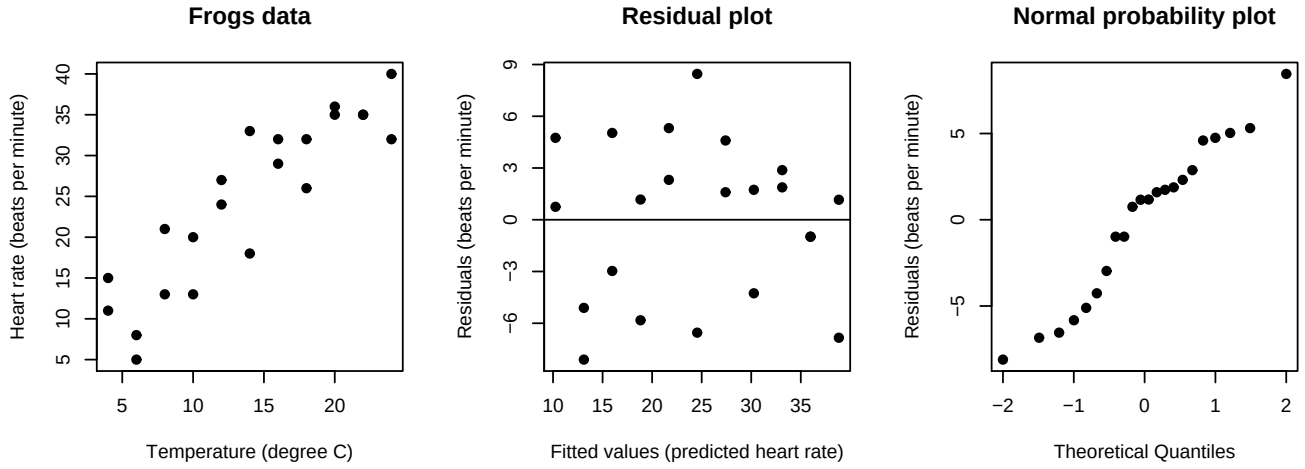
   (a) Are older or younger individuals more likely to be more satisfied with medical care? (be brief).

(b) Is the linear relationship between age and satisfaction statistically significant? Justify shortly.

(c) Below are a normal probability plot of the residuals and a plot of residuals vs. fitted values (predicted satisfaction). Under what conditions is a linear regression analysis valid? Are these conditions met here?



4. An experimenter was interested in the relationship between temperature and heart rate in the common grass frog. The temperature was manipulated in 2-degree increments ranging from 4 to 24 degrees C with heart rates recorded at each temperature in beats per minute. Two observations were made at selected each temperature, for a total of 22 observations. A different randomly selected frog was used for each observation. A scatter-plot of the data (Y = heart rate versus X = temperature) is given below along with a residual plot and a normal probability plot of the residuals. Calculations yield $\bar{x} = 14$, $\bar{y} = 24.55$, $\sum(x_i - \bar{x})(y_i - \bar{y}) = 1258$, $\sum(x_i - \bar{x})^2 = 880$, SS(total) $= \sum(y_i - \bar{y})^2 = 2237$ and SS(residual) $= 439$.



(a) Calculate the intercept and the slope of the regression line of heart rate (Y) on temperature (X). Draw the regression line on the plot.

(b) What is the predicted value of a new frog's heart rate at 15 °C?

(c) Determine the residual standard deviation s_e (and show your work). It is one of ☐ 4.572 ☐ 4.685 ☐ 4.938. How many frogs in the sample have their heart rate within two s_e from the predicted value (i.e. from the regression line)?

(d) Provide a 95% confidence interval for the slope β_1. Is the slope different from 0 with statistical significance (at $\alpha = .05$)?

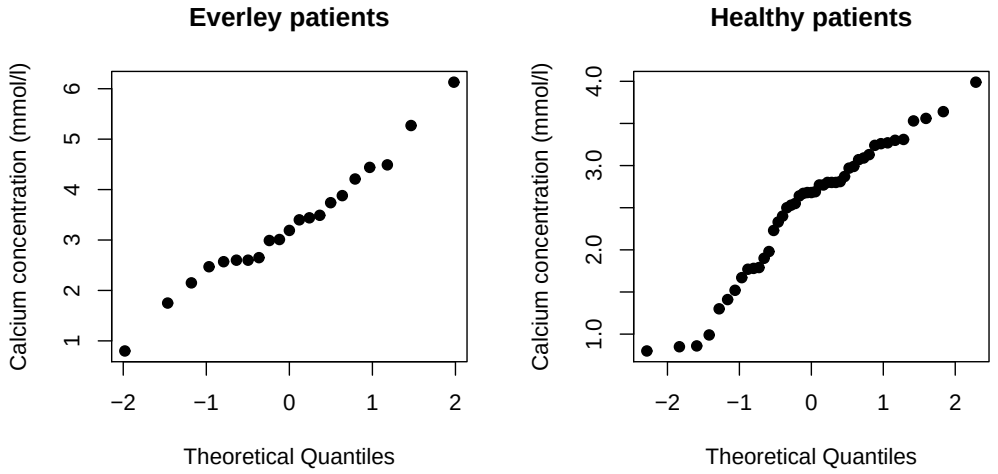
(e) Calculate the correlation coefficient between heart rate and temperature.

(f) Check whether the hypotheses underlying the regression analysis are met here.

5. A rare congenital disease, Everley's syndrome, generally causes a reduction in concentration of blood sodium. A study was conducted to determine if plasma calcium concentration is also affected in patients with Everley's syndrome. Plasma calcium concentration was measured in 21 patients with Everley's syndrome, as well as in 45 healthy patients.

The data are summarized in the following table, and normal probability plots are indicated below.

	Everley	Healthy
Mean $\bar{y}$ (in mmol/l)	3.31	2.50
SD s (in mmol/l)	1.2	0.8
n	21	45



(a) Are these data ☐ observational or ☐ experimental?

(b) Is the plasma concentration in patients with Everley's syndrome abnormally high? State first the name of the test you will use, and let $\alpha = 0.05$. Hint: the degree of freedom is one of ☐ 19.2 ☐ 28.6 ☐ 65.1

(c) Check that the method you used in 5b is valid for the data at hand.

6. A study was conducted to compare the coefficients of digestibility of dry matter for four diets fed to goats. Six randomly selected goats were assigned to each treatment. Data are indicated on the right. Sample means and standard deviations are given below. Also, calculations yield SS(error) = 250.5 and SS(total) = 689.6.

A	B	C	D
51	49	67	57
57	53	57	57
49	50	65	58
54	52	60	60
53	50	56	64
57	44	56	59

Diet	A	B	C	D
mean digestibility	53.5	49.7	60.2	59.2
SD	3.21	3.14	4.79	2.64
n	6	6	6	6

(a) Test the null hypothesis that mean coefficients of digestibility are the same for all 4 diets. State the conclusion of the test, and bracket the p-value.

(b) Is the method used in 6a valid here? (check assumptions!)

(c) Determine the pooled standard deviation s_{pooled} (show your work). It is one of
☐ 3.23 ☐ 3.30 ☐ 3.54

(d) Use the Tukey-Kramer method to compare all pairs of means at $\alpha = 0.05$. Is there sufficient evidence to tell which diet has the highest population mean digestibility, and which diet has the lowest mean digestibility? Note: in case you didn't answer question 6c, pick any value of s_{pooled} from 6c in order to carry out the Tukey-Kramer method.

Practice problems from chapters 15, 16 and 17. For earlier topics, refer to the previous list of practice problems at `www.stat.wisc.edu/courses/st371-ane/hw/practice2.pdf`.

Ch.15, p.420 & up: Practice problems 1, 2, 3, 5 (ignore d), 7 (ignore c), 8 (ignore a), 9, 10 and 12 (we have not covered non-parametric alternative to ANOVA, but you should be able to detect cases when an alternative is needed).

Ch.16, p.452 & up: Practice problems 1, 3 (ignore the d, e, and SE in c), 4, 5, 11 (for b, use a test from Chapter 17).

Ch.17, p.498 & up: Practice problems 1, 2, 3, 4, 5, 9 (ignore d-e), 10, 11 (ignore b and c; the prediction intervals are indicated by the outer dashed lines), 12 (ignore c).

Keys to past exams' problems

1. No: the 12 observations are not independent.

2. (a) $X^2 = 1.03$ $df = 2$ and $p > .20$. H_0 is accepted. (b) Design seems okay. Expected counts are ≥ 5.

3. (a) older, because slope $b_1 > 0$. (b) Yes because $p < .001$. It shows $\beta_1 > 0$. (c) all assumptions met expect that σ_e is not constant.

4. (a) $b_1 = 1.43$, $b_0 = 4.53$. (b) 26^oC. (c) 4.685. $2s_e = 9.37$ so all frogs (i.e. 22) have their heart rate within 2 s_e from the predicted value. (d) $\text{SE}_{b_1} = 0.158$ and $t = 2.086$, we get $(1.10, 1.76)$ beats/^{o}C. Yes, $\beta_1 > 0$ with statistical significance because the interval does not include 0. (e) .896 (f) everything okay

5. (a) observational (b) df = 28.6. (other possibilities can be eliminated because 19.2 is $< 21 - 1$ and 65.1 is $> 45 - 1$. Independent sample t-test, non-directional. $\text{SE}_{\bar{y}_1 - \bar{y}_2} = 0.288$, $t = 2.77$ and $.001 < p < .01$. H_0 is rejected. Yes, plasma calcium concentration is abnormally high in Everley patients. (c) everything okay. (healthy sample large enough)

6. H_0 is rejected. All diets do not have the same mean digestibility.

	df	SS	MS	F	p-value
diet	3	439.1	146.4	11.68	$.0001 < p < .001$
residual	20	250.5	12.52		
total	23	689.6			

(b) yes. (randomness, SD about the same, normal probanility plots not available, but there is no outlier) (c) 3.54 (d) scale factor 1.44.

	$\bar{y}_i - \bar{y}_2$	SE	T value	critical value	significantly different?
C vs. D	1.0	2.044	0.50	2.80	No
vs. A	6.7	2.044	3.28	2.80	Yes
vs. B	10.5	2.044	5.14	2.80	Yes
D vs. A	5.7	2.044	2.79	2.80	No but so barely, we'll say yes! (using $\alpha \approx 0.055$ or so)
vs. B	9.5	2.044	4.65	2.80	Yes
A vs. B	3.8	2.044	1.86	2.80	No
2.95	3.58	3.96			
R_i	4.263	5.174	5.723		

We get this display: <u>B A</u> <u>D C</u>

A and B are not significantly different. C and D neither. Most: either C or D (or both). Least: either A or B (or both).