MA 486/586-1C (Statistics), Dr Chernov   Final exam   Fri, Apr 25, 2014

Student's name ______________________________

1. (15 pts) Five data points are observed:

$$(2,1),\ (-3,-1),\ (3,2),\ (5,3),\ (-2,0)$$

(a) Compute the sample means $\bar{x}$ and $\bar{y}$. Answers: $\bar{x} = 1$ and $\bar{y} = 1$.
(b) Compute the sample variances s_x^2 and s_y^2. Answers: $s_x^2 = 11.5$ and $s_y^2 = 2.5$.
(c) Compute the sample covariance c_{xy} and the sample correlation coefficient r. Answers: $c_{xy} = 5.25$ and $r \approx 0.979$.
(d) Estimate the parameters α and β of the regression line $y = \alpha + \beta(x - \bar{x})$. Answers: $\alpha = 1$ and $\beta = \frac{525}{1150} \approx 0.4565$.
(e) Draw a scatter plot, mark the data points and the regression line.

2. (15 pts) A computer program supposedly generates a standard random variable $\mathcal{N}(0,1)$. The following are five numbers produced by this program:

$$0.3,\ 2.1,\ -1.71, -0.92, 1.13$$

Use the Kolmogorov-Smirnov test to test the hypothesis that the program works right.
(a) Sketch the empirical distribution function.
(b) Compute the test statistic D. Answer: $D = 0.2708$.
(c) Assume significance level 20%. Find the critical value d. Answer: d=0.45.
(d) Which hypothesis do you accept? Answer: H_0.
(e) Sketch a 95% confidence band around the empirical distribution function.

3. (15 pts) Given random sample

$$4,\ 7,\ 11,\ 2,\ -2,\ 3,\ -5,\ 6,\ 2,\ 0,\ 8,\ -7,\ 5,\ 1,\ -1$$

(a) Find a point estimate for the median, m (write the formula you use). Answer: $\hat{m} = y_8 = 2$.
(b) Find point estimates for the quartiles, $\pi_{0.25}$ and $\pi_{0.75}$ (write the formulas you use). Answers: $\hat{\pi}_{0.25} = y_4 = -1$ and $\hat{\pi}_{0.75} = y_{12} = 6$.
(c) Find an approximate 96% confidence interval for the median, m.
Give the exact confidence level. Answers: $(y_4, y_{12}) = (-1, 6)$; level=0.9648.
(d) Find $\mathbb{P}(1 < \pi_{0.7} \leq 7)$ (write the formulas you use). Answer: $\mathbb{P}(6 \leq b(15, 0.7) \leq 12) = 0.8695$.

4. (15 pts) In a college, students are classified according to the period of the day when they took calculus (morning, afternoon, evening) and the success or failure in the course:

	Morning	Afternoon	Evening
Pass	80	60	10
Fail	28	20	2

Test the hypothesis that the success rate is independent of the time period of the class.
(a) Estimate the probabilities p_{ij}. Answers: 0.54, 0.4, 0.06.
(b) Find the theoretical frequencies $n\hat{p}_{ij}$. Answers: 81, 60, 9, 27, 20, 3.
(c) Compute the test statistic Q (as the sum of six terms). Answer: $Q = 0.4938$.
(d) How many degrees of freedom are here? Answer: 2.
(e) Assume significance level 10%. What is the critical value? Answer: $\chi^2_{0.1}(2) = 4.605$.
(f) Which hypothesis do you accept? Answer: H_0.

5. (10 pts) Let X_1, X_2, X_3 be three independent random variables that have some distributions with mean values μ_1, μ_2, μ_3 respectively. Test the hypothesis

$$H_0 : \mu_1 = \mu_2 = \mu_3$$

at 1% significance level. The observed data are given in the table below:

X_1 :	8	6	7			
X_2 :	3	4	6	4	5	8
X_3 :	10	8	11	11		

(a) Compute $\bar{x}_{i\cdot}$ for $i = 1, 2, 3$ and $\bar{x}_{\cdot\cdot}$. Answers: 7,5,10, and 7.
(b) Compute SS(E) and SS(T). Answers: SS(E)= 24 and SS(T)= 60.
(c) Compute the F statistic. Answer: $F = 12.5$.
(d) What is the critical value? Answer: $F_{0.01}(2, 10) = 7.56$.
(e) Which hypothesis do you accept? Answer: H_1.

6. (10 pts) Two samples from random variables X and Y were recorded:

$$X: \ 14, \ 11, \ 3, \ 12, \ 2, \ 11, \ 5, \ 17, \ 12, \ 4$$

$$Y: \ 6, \ 18, \ -2, \ 7, \ 0, \ 19, \ -5, \ 9, \ 6, \ 20$$

Use the Wilcoxon test (for two samples) to test the hypothesis $H_0 : m_X = m_Y$ against $H_1 : m_X \neq m_Y$.

(a) Compute W. Answer: $W = 101$.
(b) Find μ and σ^2. Answers: $\mu = 105$ and $\sigma^2 = 175$.
(c) Compute the Z-score and the p-value. Answer: $Z = -0.3024$ and p-value$= 0.7624$.
(d) Which hypothesis would you accept? Answer: H_0.

7. (10 pts) Two samples from random variables X and Y were observed:

$$X: \ 14, \ 11, \ 3, \ 12, \ 2, \ 11, \ 5, \ 17, \ 12, \ 4$$

$$Y: \ 6, \ 18, \ -2, \ 7, \ 0, \ 19, \ -5, \ 9, \ 6, \ 20$$

Use the run test to test the hypothesis $H_0 : F_X = F_Y$ against $H_1 : F_X \neq F_Y$.

(a) Compute R. Answer: $R = 5$.
(b) Find μ and σ^2. Answers: $\mu = 11$ and $\sigma^2 = \frac{90}{19} \approx 4.737$.
(c) Compute the Z-score and the p-value. Answer: $Z = -2.757$ and p-value$= 0.0029$.
(d) Which hypothesis would you accept? Answer: H_1.

8. (10 pts) Estimate the main effect, the three two-factor interactions, and the three-factor interaction in a 2^3 factorial design experiment. The data, in the canonical order (see Section 20.6), are:

$$x_1 = 4, \quad x_2 = 2, \quad x_3 = 3, \quad x_4 = 1, \quad x_5 = 5, \quad x_6 = 3, \quad x_7 = 6, \quad x_8 = 4.$$

(a) Compute [A], [B], [C], [AB], [AC], [BC], and [ABC]. Answers: -1, 0, 1, 0, 0, 0.5, 0.
(b) Construct an approximate $q - q$ plot.
(c) Comment on it.

[Bonus, up to 20 pts] Suppose in a sequence of 100 independent trials 20 successes (and 80 failures) are observed. Let p denote the unknown probability of success.

(a) Give the maximum likelihood estimate $\hat{p}$ for p. Answer: 0.2.

(b) Find the bias and give a formula for the variance of your estimate. Answers: the bias is zero, and $\mathsf{Var}\hat{p} = \frac{p(1-p)}{100}$.

(c) Construct a 98% confidence interval for p. Answer: 0.2 ± 0.093.

(d) Test hypothesis $H_0: \; p = 0.3$ against $H_1: \; p < 0.3$ at 2% significance level. Answer: $Z = -2.182$, critical value is $-z_{0.02} = -2.054$, we accept H_1.

(e) How many trials need to be performed so that the margin of error in a 98% confidence interval for p be less than 0.02? Answer: 3382.