University of California, Los Angeles
Department of Statistics

Statistics 13 — **Instructor: Nicolas Christou**

Exam 2
21 November 2014

Name: SOLUTIONS

UCLA ID: ____________________

Section: ______

Problem 1 (25 points)
Answer the following questions:

a. A missile protection system consists of n radar sets operating independently, each with probability 0.90 of detecting a missile entering a zone. If $n = 5$ and a missile enters the zone what is the probability that exactly 4 radar sets detect the missile?

$X \sim b(5, 0.90)$ $\quad P(X=4) = \binom{5}{4} 0.90^4 \, 0.10^1 = 0.32805$.

b. Refer to question (a). How large must n be if we require that the probability of detecting a missile that enters the zone be 0.999?

$P(\text{AT LEAST ONE SUCCESS}) = 0.999$ OR $P(X \geq 1) = 0.999$

OR $1 - P(X=0) = 0.999$ OR $1 - 0.1^n = 0.999 \Rightarrow n = \frac{\log(0.001)}{\log(0.1)} = 3$.

c. A true-false test consists of 50 questions. How many does a student have to get right to convince you that he/she is not just guessing? Please explain.

$\mu = np = 50 \frac{1}{2} = 25$

$\sigma = \sqrt{np(1-p)} = \sqrt{50 \frac{1}{2} \frac{1}{2}} = 3.54$

MUST GET $\mu + 2.5(\sigma)$ OR MORE

$25 + 2.5(3.54) \simeq 34$

33–34 OR MORE.

d. Only 4% of people have Type AB blood. Suppose donors arrive independently at the UCLA Blood and Platelet Center. What is the probability that we find the 3rd Type AB blood donor on the 8th trial?

$\underbrace{\_\ \_\ \_\ \_\ \_\ \_\ \_}_{\text{NEED 2 SUCCESSES AMONG 7 TRIALS}} \mid$ 3RD SUCCESS ON 8TH TRIAL

$\binom{7}{2} 0.04^2 \, 0.96^5 \times 0.04 = 0.0011$

e. Suppose blood donors arrive at the UCLA Blood and Platelet Center according to the Poisson distribution at the rate of $\lambda = 10$ per hour. What is the probability that in the next 30 minutes at least two blood donors will arrive?

$\lambda = 5 \quad P(X \geq 2) = 1 - [P(X=0) + P(X=1)]$

$= 1 - \left[\frac{5^0 e^{-5}}{0!} + \frac{5^1 e^{-5}}{1!}\right] = 0.9596$.

Problem 2 (25 points)
Answer the following questions:

a. The acidity of soils is measured by pH, which may range from 0 (high acidity) to 14 (high alkalinity). A soil scientist wants to estimate the average pH for a large field by randomly selecting $n = 40$ samples. Suppose $\sigma = 0.75$. Find the probability that the sample mean of the 40 pH measurements will be within 0.2 unit of the true mean pH of the field.

$$P\left[\mu-0.2 < \bar{X} < \mu+0.2\right] = P\left[\frac{\mu-0.2-\mu}{0.75/\sqrt{40}} < Z < \frac{\mu+0.2-\mu}{0.75/\sqrt{40}}\right] = P\left(-1.69 < Z < 1.69\right)$$

$$= 0.9545 - 0.0455 = 0.9090.$$

b. Refer to question (a). Suppose the scientist would like the sample mean to be within 0.1 of the true mean with probability 90%. How many sample should the scientist take?

$$P\left[\mu-0.1 < \bar{X} < \mu+0.1\right] = 0.90 \Rightarrow P\left(\frac{\mu-0.1-\mu}{0.75/\sqrt{n}} < Z < \frac{\mu+0.1-\mu}{0.75/\sqrt{n}}\right) = 0.90$$

$$\Rightarrow P\left[-\frac{0.1\sqrt{n}}{0.75} < Z < \frac{0.1\sqrt{n}}{0.75}\right] = 0.90 \rightarrow 1.645 = \frac{0.1\sqrt{n}}{0.75} \Rightarrow n \approx 153$$

c. We discussed in class that when two independent large samples are selected, the difference in the sample means $(\bar{X} - \bar{Y})$ according to the central limit theorem follows approximately the normal distribution. The flow of water through soil depends on, among other things, the porosity of the soil. For the comparison of two types of sandy soil, $n_1 = 50$ measurements and $n_2 = 100$ measurements are to be taken from two populations. Assume that $\sigma_1^2 = 0.01$ and $\sigma_2^2 = 0.02$. Find the probability that the difference between the sample means will be within 0.05 unit of the difference between the population means μ_1 and μ_2.

$$P\left[\mu_1-\mu_2-0.05 < \bar{X}-\bar{Y} < \mu_1-\mu_2+0.05\right] = P\left[\frac{\mu_1-\mu_2-0.05-(\mu_1-\mu_2)}{0.02} < Z < \frac{\mu_1-\mu_2+0.05-(\mu_1-\mu_2)}{0.02}\right]$$

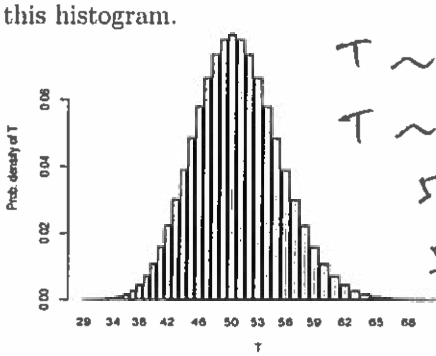
$$\bar{X}-\bar{Y} \sim N\left(\mu_1-\mu_2, \sqrt{\frac{0.01}{50}+\frac{0.02}{100}}\right)$$

OR $\bar{X}-\bar{Y} \sim N(\mu_1-\mu_2, 0.02)$

$$= P\left[-2.5 < Z < 2.5\right)$$

$$= 0.9938 - 0.0062 = 0.9876.$$

d. A population has mean $\mu = 1.25$ and standard deviation $\sigma = 2$. Repeated samples each one of size $n = 40$ are selected from this population. It is claimed that the histogram below represents the distribution of the total (T) of these 40 observations. Clearly explain if there is anything wrong with this histogram.



$T \sim N(n\mu, \sigma\sqrt{n})$

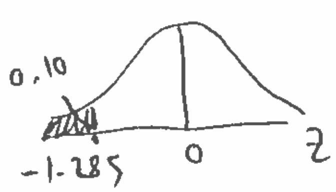
$T \sim N(50, 12.65)$

$50 \pm 3(12.65)$

50 ± 38 OR $(12, 88)$

HISTOGRAM IS TOO NARROW.

e. Refer to question (d). Find the 10th percentile of the distribution of T.



$$-1.285 = \frac{T-50}{12.65} \Rightarrow T = 33.74.$$

Problem 3 (25 points)

Normal temperature (degrees Fahrenheit) readings (x) and heart rates (beats per minute) of 32 males were used in this problem, The regression output of y on x is given below. Note: Several values are missing! You are also given: $\sum_{i=1}^{n} x_i = 3142.4$, $var(x) = 0.4077$, $\bar{y} = 74.5313$, $var(y) = 37.2$ and $r = 0.2194$.

```
> q <- lm(a$rate ~ a$temperature)
> summary(q)

Coefficients:
               Estimate   Std. Error  t value  Pr(>|t|)
(Intercept)                  167.135   -0.786
a$temperature
```

Answer the following questions:

a. Find $\hat{\beta}_0$.

b. Find $\hat{\beta}_1$.

c. Is there a linear association between rate and temperature? Use $\alpha = 5\%$.

d. Approximate the p-value using your t table for testing $H_0 : \beta_0 = 0$ against the alternative $H_0 : \beta_0 \neq 0$.

e. Please list few nice properties about the least squares estimates $\hat{\beta}_0$ and $\hat{\beta}_1$.

Problem 4 **(25 points)**
Answer the following questions:

a. Human skin can be removed from a dead person, stored in a skin bank, and then grafted onto burn victims. Because persons with infections, cancer, and so on must be excluded only 7% of all potential donors are acceptable. Suppose that at a large hospital they screen 8000 deaths and from these they get only 80 donors. Do the data provide sufficient evidence that this hospital is selecting its donors from a population containing a lower percentage of acceptable donors than found in the population at large (7%)? List all the steps using the theory-based test.

$H_0: P = 0.07$

$H_a: P < 0.07$

$$z = \frac{\hat{p} - P_0}{\sqrt{\frac{P_0(1-P_0)}{n}}} = \frac{0.01 - 0.07}{\sqrt{\frac{0.07(1-0.07)}{8000}}} = -21.03$$

$\hat{p} = \frac{80}{8000} = 0.01$

REJECT Ho.

b. Explain clearly how you would test the hypothesis of question (a) using simulations.

BOOTSTRAP CONFIDENCE INTERVALS.

c. In hypothesis testing what parameters you think affect the power of the test and in what way? Note: List all the parameters that affect the power of the test and indicate whether the power will increase, decrease, or stay the same.

	POWER $(1-\beta)$
α ↓	↓
SD ↓	↑
n ↑	↑
SHIFT ↑	↑

d. Is it true that in Poisson distribution the standard deviation is always less than the mean? How about in the binomial distribution?

POISSON

$EX = \lambda$

$VAR(X) = \lambda$ YES.

$SD(X) = \sqrt{\lambda}$

BINOMIAL

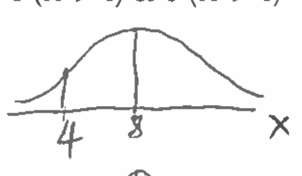
$EX = np$

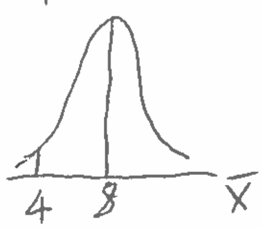
$VAR(X) = np(1-p)$ NO.

$SD(X) = \sqrt{np(1-p)}$ IT DEPENDS.

e. If $\bar{X}$ is the sample mean of n observations selected from $X \sim N(8, \sigma)$ explain which one is larger $P(\bar{X} > 4)$ or $P(X > 4)$

$X \sim N(8, \sigma) \rightarrow \bar{X} \sim N\left(8, \frac{\sigma}{\sqrt{n}}\right)$





$P(\bar{X} > 4)$ IS LARGER.