PROBABILITY AND STATISTICS FOR ENGINEERS MIDTERM 2		
Code : *CVE 303* Acad.Year : *2019-20* Semester : *Fall* Date : *30.11.2019* Time : *10:40* Duration : *110 min*	Last Name : Name : Student ID : Signature : 7 QUESTIONS ON 4 PAGES TOTAL 100 POINTS	#:

P1. (28)	P2. (24)	P3. (28)	P4. (20)	B. (10)	Total. (100)

1. *(8×3=24pts)* Short answer questions: Point Estimation.

(A) If qnorm(x) = 0.5, then what is x?

$x = \text{pnorm}(0.5)$

$P(Z < 0.5) =$ probability that normal random variable has value $< \frac{1}{2} \cdot$(standard deviation)

(B) What is the difference between "sample mean" and "population mean"?

sample mean $= \bar{x}$
$=$ average of measured values

population mean $= \mu$
$=$ average of distribution

(C) What does it mean for an estimator to be "biased"?

"Biased" estimator has wrong expected value.

$E[\hat{\Theta}] \neq \theta$

(D) What two properties should a good estimator have?

(1) Unbiased $(E[\hat{\Theta}] = \theta)$

(2) Small Variance (Standard Error $\sigma_{\hat{\theta}}$ small)

(E) Give a formula for the standard error of the sample mean.

$$\sigma_{\bar{x}} = \sqrt{\sigma^2/n} = \sigma/\sqrt{n}$$

[• $\sigma^2 = Var[X]$
• $n =$ # samples]

(F) Give a formula for the standard error of the sample population proportion.

$$\sigma_{\hat{p}} = \sqrt{pq/n}$$

[• $p =$ population proportion
• $q = (1-p)$]

(G) Name two unbiased estimators for $\mu = \mathrm{E}[X]$.

$\bar{X} =$ sample mean

$\bar{X}_{tr} =$ trimmed sample mean

$\tilde{X} =$ sample median

⋮

(H) What is a "point estimator"?

$\hat{\Theta}$ a random variable used to estimate the value of a parameter θ

– usually a statistic of multiple IID sample random variables

2. *(6×2=12pts)* Short answer questions: Confidence intervals.

(A) What is a $(1-\alpha)\%$ confidence interval?

An interval I around a point estimate $\hat{\theta}$ for a parameter θ so that $P(\theta \in I) = 1 - \alpha$

(Equivalently $P(\theta \notin I) = \alpha$)

(B) Is it better for confidence intervals to be wide or narrow?

Narrow confidence intervals are better

– more accurate point estimation.

(C) How can you change the width of a confidence interval?

$w = 2 \frac{\sigma}{\sqrt{n}} \left| \underbrace{\text{qnorm}(\alpha/2)}_{z_{\alpha/2}} \right|$

Increase n $\Rightarrow$ Decrease w
Decrease σ $\Rightarrow$ Decrease w
Increase α $\Rightarrow$ Decrease w

(D) When do you use χ^2-distribution for confidence intervals?

When approximating variance σ^2
(or std. dev. σ)

(E) What is a prediction interval?

An interval around a point estimate $\hat{\theta}$ where the next sample value will be in interval with given probability

$P(x_{n+1} \in I) = 1 - \alpha \qquad (I = \hat{\theta} \pm E_{\alpha/2})$

(F) What is the standard error of $(\overline{X} - X)$?

$\sigma_{\overline{X}}^2 = \sigma^2/n$ and $\sigma_X^2 = \sigma^2$

so $\sigma_{\overline{X}-X}^2 = \sigma^2/n + \sigma^2$

std error $= \sigma_{\overline{X}-X} = \sqrt{\sigma^2/n + \sigma^2}$
$= \sigma \sqrt{1/n + 1}$

3. *(8×2=16pts)* Short answer questions: Hypothesis testing.

(A) In hypothesis testing what probability is the p-value giving?

$p\text{-value} = P\left(\text{data as extreme as observed} \,\middle|\, H_0 : \theta = \theta_0\right)$

"Probability of observation if H_0 true"

(B) Is it bad to reject the null hypothesis?

No. The goal of hypothesis test is to reject null hypothesis!

(C) In hypothesis testing do you want the p-value to be big or small?

Small. Smaller p-value means more likely to reject null hypothesis!

(D) In hypothesis testing do you want the score to be big or small?

Big. Big score means observation is many std. errors away from H_0 so p-value is small!

(E) In hypothesis testing, what probability is α giving?

$\alpha = P(\text{Type I Error})$
$= P(\text{Reject } H_0 \mid H_0 \text{ True})$

α is probability cutoff for rejecting H_0

(F) What is a "Type II Error" in hypothesis testing?

"Type II Error" is Fail to reject H_0 when H_0 false.

(G) List two unethical (bad) ways to change p-value.

- Reduce s by trimming extreme samples
- Rerunning experiment without accounting for old data
- Arbitrarily changing to 1-tailed test

(H) Give one ethical (good) way to change p-value.

- Increase n (#samples)

4. *(7×2=14pts)* Suppose X is sampled 81 times, with a sample mean of 21 and a sample variance of 4. (Your answers should be in terms of the R commands `pnorm(..)`, `qnorm(..)`, `pt(..)`, `qt(..)`, etc)

(A) Find the (2-sided) 90% confidence interval for $\mu = \mathrm{E}[X]$.

$n = 81$ n is big, so we use Normal distribution

$\bar{x} = 21$

$s^2 = 4$ $Z = \frac{\bar{x}-\mu}{s/\sqrt{n}} \sim \mathrm{Normal}(0,1)$

$\hookrightarrow s = \sqrt{4} = 2$

$s/\sqrt{n} = 2/\sqrt{81} = 2/9$

$\mu \approx \bar{x} \pm s/\sqrt{n}\ \mathrm{qnorm}(\alpha/2)$

$21 \pm 2/9\ \mathrm{qnorm}(.05)$

(B) What is the (2-tailed) p-value for $\mathrm{H}_0: \ \mu = 23$?

$n = 81$ n is big so we use z-test (Normal)

$\bar{x} = 21$

$s^2 = 4$ $Z = \frac{\bar{x}-\mu_0}{s/\sqrt{n}} \sim \mathrm{Normal}(0,1)$

$p = 2 \cdot \mathrm{pnorm}\left(-\left|\frac{\bar{x}-\mu_0}{s/\sqrt{n}}\right|\right)$

$= 2\,\mathrm{pnorm}\left(-\frac{2}{2/9}\right) = 2\,\mathrm{pnorm}(-9)$

(a very small number...)

5. *(7×2=14pts)* Suppose X is sampled 9 times, with a sample mean of 21 and a sample variance of 4. (Assume $X \sim$ Normal.)

(A) Find the (2-sided) 90% confidence interval for $\mu = \mathrm{E}[X]$.

$n = 9$ n is small so we use t-distribution

$\bar{x} = 21$ (degrees of freedom $= n-1 = 8$)

$s^2 = 4$

$\hookrightarrow s/\sqrt{n} = 2/3$ $t = \frac{\bar{x}-\mu}{s/\sqrt{n}} \sim t(n-1) = t(8)$

$\mu \approx \bar{x} \pm s/\sqrt{n}\ \mathrm{qt}(\alpha/2, n-1)$

$21 \pm 2/3\ \mathrm{qt}(.05, 8)$

(B) What is the (2-tailed) p-value for $\mathrm{H}_0: \ \mu = 23$?

$n = 9$ n is small so we use t-test

$\bar{x} = 21$

$s^2 = 4$ $t = \frac{\bar{x}-\mu_0}{s/\sqrt{n}} \sim t(n-1) = t(8)$

$p = 2 \cdot \mathrm{pt}\left(-\left|\frac{\bar{x}-\mu_0}{s/\sqrt{n}}\right|, n-1\right)$

$= 2\,\mathrm{pt}(-3, 8)$

6. *(7+6+7=20pts)* Suppose 90 samples are gathered, finding 30 defectives.

(A) Find the (2-sided) 90% confidence interval for the proportion of defectives.

$n = 90$

$x = 30$

$\hat{p} = x/n = 1/3$

$\hookrightarrow \hat{q} = 1 - \hat{p} = 2/3$

$$\frac{\hat{p} - p}{\sqrt{pq/n}} \approx \text{Normal}(0,1)$$

$\hookrightarrow$ use $\sqrt{\hat{p}\hat{q}/n} = \sqrt{\frac{1/3 \cdot 2/3}{90}} = \frac{1}{9\sqrt{5}}$

$$p \approx \hat{p} \pm \sqrt{\hat{p}\hat{q}/n}\ \text{qnorm}(\alpha/2)$$

$$1/3 \pm \frac{1}{9\sqrt{5}}\ \text{qnorm}(.05)$$

(B) Find the (2-sided) 90% confidence interval for the number defectives in a population of 900.

multiply by 900

$$E[\#\text{defectives in population of } 900] = p \cdot 900$$

$$p \cdot 900 \approx \left[\hat{p} \pm \sqrt{\hat{p}\hat{q}/n}\ \text{qnorm}(\alpha/2)\right] \cdot 900$$

$$\left[1/3 \pm \frac{1}{9\sqrt{5}}\ \text{qnorm}(.05)\right] \cdot 900$$

$$300 \pm 20\sqrt{5}\ \text{qnorm}(.05)$$

(C) What is the (2-tailed) p-value for $\quad H_0: \ p = 1/6$?

$n = 90$

$x = 30$

$\hat{p} = x/n = 1/3$

$$\frac{\hat{p} - p}{\sqrt{pq/n}}$$

$\hookrightarrow$ use $\sqrt{p_0 q_0/n} = \sqrt{\frac{1/6 \cdot 5/6}{90}} = \frac{1}{18\sqrt{2}}$

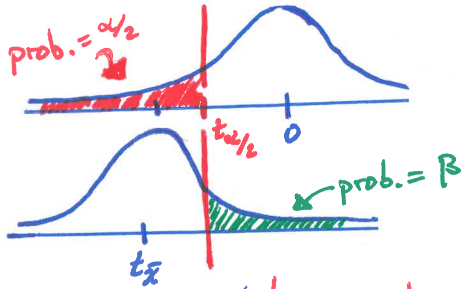
$$z = \frac{\hat{p} - p_0}{\sqrt{p_0 q_0/n}} = \frac{1/3 - 1/6}{1/18\sqrt{2}} = 3\sqrt{2}$$

$$p = 2 \cdot \text{pnorm}(-3\sqrt{2})$$

≈ 2(a very small number...)

Note: We only used Binomial for 1-tailed test... 2-tailed is more complicated because Binomial is not symmetric

BONUS *(2×5=10 Bonus)* Suppose you expect data to have $\bar{x} \approx 10$ and $s^2 \approx 4$ and you wish to test against $H_0: \ \mu = 12$ with significance level $\alpha = 1\%$.



(A) If $n = 16$ then what is the power of the test?

$\hookrightarrow$ t-distribution!

$$t_{\bar{x}} = \frac{10 - 12}{\sqrt{4/16}} = -\frac{2}{1/2} = -4$$

$$t_{\alpha/2} = \text{qt}(.005, 15)$$

$\hookrightarrow$ two-sided t-test with $\alpha/2 = .1/2 = .05$

$$\beta = \text{pt}(-|t_{\bar{x}} - t_{\alpha/2}|, 15)$$

$$\text{Power} = 1 - \beta$$

(B) If you wish to have a test with power 90% then how many samples are required?

(Assume n will be big enough to use normal distribution)

$z_{\alpha/2} = \text{qnorm}(.005)$

$z_\beta = \text{qnorm}(.1)$

$$n = \left[\frac{s}{\bar{x} - \mu_0}(z_{\alpha/2} + z_\beta)\right]^2 = \left[\frac{4}{12 - 10}(z_{\alpha/2} + z_\beta)\right]^2$$

$$= 4\,(z_{\alpha/2} + z_\beta)^2$$