# STATISTICS 641 - Exam 3

**Total time is 120 minutes. The exam is available for a 24 hours window.**

Name ______________________________________________

Email Address ______________________________________

Please put your answers in the following table.

| 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|
| | | | | |
| 6 | 7 | 8 | 9 | 10 |
| | | | | |
| 11 | 12 | 13 | 14 | 15 |
| | | | | |
| 16 | 17 | 18 | 19 | 20 |
| | | | | |
| 21 | 22 | 23 | 24 | 25 |
| | | | | |

Please TYPE your name and email address, and **ANSWERS in the above table. This answer table will be graded.** Often we have difficulty in reading the handwritten names and email addresses. Make this cover sheet the first page of your Solutions.

**STATISTICS 641 - Exam #3**

It is claimed that the lifetime of a newly designed battery in electronic cars follows an exponential distribution with mean $m$ years. Suppose that 5 such batteries are randomly chosen, and their lifetimes are recorded as $X_1, \cdots, X_5$. The sample average and sample variance of these 5 observations are denoted by $\bar{X}$ and $S^2$. Answer the next four questions.

1. Find the probability that a randomly chosen battery will not survive more than 3 years? Assume that $m = 7.2$.

   (a) 0.3407594 •
   (b) 0.8703247
   (c) 0.1296753
   (d) 0.1703797

   Here lifetime $X \sim$ Exponential such that mean is 7.2, therefore,

   $$f(x) = \frac{}{7.2} \exp(-x/7.2).$$

   pr(will not survive more than 3 years) $=$ pr$(X \leq 3) = 1 - \exp(-3/7.2) = 0.3407$.

2. What is the distribution of the average lifetime of these 5 batteries? Use $m = 7.2$.

   (a) Exponential with mean 1.44
   (b) Gamma with shape and scale 5 and 1.44, respectively •
   (c) Normal with mean 1.44 and varinace 10.368
   (d) Normal with mean 7.2 and varinace 10.368
   (e) Exponential with mean 7.2

   Due to additive property $\sum_{i=1}^{5} X_i$ will follow Gamma$(5, 7.2)$. The average $\sum_{i=1}^{5} X_i/5$ will follow Gamma$(5, 7.2/5)$.

3. In order to know the variability of the lifetime distribution we need to estimate $m^2$, where $m$ =mean of the lifetime distribution. An unbiased estimator of $m^2$ based on $X_1, \cdots, X_5$ is

   (a) $\sqrt{S^2}$
   (b) $(10)^{-1} \sum_{i=1}^{5} X_i^2$ •
   (c) $\bar{X}^2$
   (d) $\sum_{i=1}^{5} X_i^2/5$
   (e) $0.5\{S^2 + \bar{X}^2\}$

   Observe that

   $$\begin{aligned} E(X^2) &= \int_0^\infty x^2 \frac{1}{m} \exp(-\frac{x}{m})dx \\ &= (m^2) \int_0^\infty \left(\frac{x}{m}\right)^2 \frac{1}{m} \exp(-\frac{x}{m})dx \\ &= m^2 \int_0^\infty z^2 \exp(-z)dz = (m)^2\Gamma(3) = 2 \times (m)^2. \end{aligned}$$

   Thus $E(\sum_{i=1}^{5} X_i^2) = 5E(X^2) = 5 \times 2 \times m^2$. Thus, $10^{-1}E(\sum_{i=1}^{5} X_i^2) = m^2$.

4. Suppose that $\bar{X} = 6.6$ years. A 95% confidence interval for $m$ is [you may use pr$(\chi^2_4 < 0.4844) = 0.025$, pr$(\chi^2_4 < 11.1433) = 0.975$, pr$(\chi^2_5 < .8312) = 0.025$, pr$(\chi^2_5 < 12.8325) = 0.975$, pr$(\chi^2_{10} < 3.2469) = 0.025$, pr$(\chi^2_{10} < 20.4832) = 0.975$]

   (a) $(2.572, 39.701)$
   (b) $(5.143, 79.402)$
   (c) $(3.222, 20.327)$ •
   (d) $(2.961, 68.123)$

(e) $(0.592, 13.624)$

The formula is (page 10, HO 11)

$$\left(\frac{2n\bar{X}}{\chi^2_{2n,1-\alpha/2}}, \frac{2n\bar{X}}{\chi^2_{2n,\alpha/2}}\right).$$

Hence,

```
10*6.6/qchisq(0.025, 10)
[1] 20.32663
> 10*6.6/qchisq(0.975, 10)
[1] 3.222156
```

5. Suppose that we know $0 \leq p \leq 0.15$. Find $n$ such that we are 99% confident that $\widehat{p}$ is within 0.01 units of $p$ (use Wald's formula, and choose the best possible answer).

   (a) 8461 •

   (b) 5973

   (c) 4899

   (d) 3458

   (e) 10622

   The formula is $Z_{\alpha/2}p(1-p)/d^2$. If you know $a < p < b$, replace $p$ by the value which is closest to 0.5. In our case $0 < p < 0.15$, so replace $p$ by 0.15.

```
 qnorm(0.005)^2*0.15*(1-0.15)/0.01^2
= 8459.493
```

In an extrasensory perception expreiment (ESP), five choices were offered for each question. Assume that a person without ESP gusses and thus correctly answers with probability (1/5). Assume that responses are independent from question to question, and altogether 100 questions were asked. Answer the next two questions.

6. What are the mean and standard deviation of the proportion of correct answers?

   (a) 0.2 and 0.04 •
   (b) 0.2 and 0.4
   (c) 0.002 and 0.04
   (d) 0.4 and 0.02
   (e) 5 and 0.4

   From your HW. $E(\hat{p}) = p = 1/5 = 0.2$ and $\text{var}(\hat{p}) = p(1-p)/n = 0.2(1-0.2)/100$. So, the standard error is $\sqrt{0.2 \times 0.8/100} = 0.04$.

7. What is the probability that a person without ESP will correctly answer at least 30 questions out of 100 questions (use normal approximation, and use $1 - \text{pr}(X \leq 29)$).

   (a) 0.2156
   (b) 0.065
   (c) 0.49389
   (d) 0.022
   (e) 0.01222 •

   Let $X$ be the number of correct answers out of 100. Then actually $X \sim \text{Binomial}(100, 0.2)$, and approximately $X \sim \text{Normal}(20, 4^2)$ To calculate $\text{pr}(X \geq 30) = 1 - \text{pr}(X \leq 29) = 1 - \Phi\{(29-20)/4\} = 0.01222447$.

8. Suppose that $Y_1, \cdots, Y_{10}$ are iid observations from a normal distribution. The prediction interval for a future observation will be

   (a) $\bar{Y} \pm t_{\alpha/2,n-1} S\sqrt{(n-1)/n}$
   (b) $\bar{Y} \pm t_{\alpha/2,n-1} S\sqrt{1+\frac{1}{n}}$ •
   (c) $\bar{Y} \pm t_{\alpha/2,n-1} S/\sqrt{n}$
   (d) $\bar{Y} \pm t_{\alpha/2,n} S\sqrt{1+\frac{1}{n}}$
   (e) None of the above is correct as the sample size is small.

   Page 36, HO 11.

   Suppose that for testing $H_0 : \mu = 2$ vs $H_a : \mu > 2$ for a normal population with $\sigma = 1$, 16 observations are drawn at random. Assume that the rule is that we reject $H_0$ when $\bar{X} > 2.5$. Answer the next three questions.

9. What is the probability of Type-I error?

   (a) 0.0006
   (b) 0.0227 •
   (c) 0.0014
   (d) 0.1586
   (e) 0.0793

   $\text{pr(type-I error)} = \text{pr}(\text{Reject } H_0|\mu = 2) = \text{pr}(\bar{X} > 2.5|\mu = 2)$. Note that when $\mu = 2$, $\bar{X} \sim \text{Normal}(2, \sigma^2_{\bar{X}} = 1/16)$. Thus $\text{pr}(\bar{X} > 2.5|\mu = 2) = 1 - \Phi\{(2.5-2) \times 4\} = 0.0227$.

10. What is the probability of Type-II error when $\mu = 2.5$?

    (a) 0.0454
    (b) 0.5 •
    (c) 0.0227
    (d) 0.7565
    (e) 0.08

pr(type-II error) = pr(fai to reject $H_0|\mu = 2.5$) = pr($\bar{X} < 2.5|\mu = 2.5$) = $\Phi\{(2.5 - 2.5) \times 4\} = \Phi(0) = 0.5$.

11. Suppose the observed value of $\bar{X} = 2.45$. What is the p-value of the test?

    (a) 0.0144
    (b) 0.5792
    (c) 0.0525
    (d) 0.4207
    (e) 0.0359 •

    p-value= pr($\bar{X} > 2.45|H_0$ is true) = pr($\bar{X} > 2.45|\mu = 2$) = $1 - \Phi\{(2.45 - 2) \times 4\} = 0.0359$.

12. What sample size is necessary if we want to reject $H_0 : \mu = 2$ against $H_a : \mu > 2$ with 80% probability when $\mu = 2.2$ along with probability of type-I error 0.01? Use $\sigma = 1$.

    (a) 344
    (b) 251 •
    (c) 292
    (d) 76
    (e) 36

    page 18, HO 12, $n = \{\sigma(Z_\alpha + Z_\beta)/\delta\}^2$. Here $\delta = 2.2 - 2 = 0.2$, $\sigma = 1$, $Z_\alpha = Z_{0.01} = 2.326$, $\beta$ = pr(type-II error) = $1 -$ power $= 1 - 0.80 = 0.20$. So, $Z_\beta = Z_{0.20} = 0.8416$. Hence we get $n = 250.9$.

13. Which of the following statements is correct regarding bootstrap precedure?

    (a) The simulation error is controlled by the number of bootstrap samples $B$. •
    (b) The statistical error is controlled by the number of bootstrap samples $B$.
    (c) The statistical error arises due to difference between the true population distribution and simulated distrubution based on $B$ bootstrap samples.
    (d) The statistical error can be greatly reduced if we use not-equal probability sample instead of simple random sample.
    (e) While the statistical error remain unchanged, the simulation error is reduced by one half if the number of observations is doubled.

    Page 71, HO 11

14. Which of the following statements is correct regarding hypothesis testing?

    (a) Calculation of p-value requires the knowledge of the test statistic's distribution under $H_a$.
    (b) Rejecting $H_0$ when in fact $H_0$ is true is the type-I error. •
    (c) Probability of Type-II error is the same as the level of significance $\alpha$.
    (d) Power of a test is the probability of type-II error.
    (e) The distribution of a test statistic is exactly the same under $H_0$ and $H_a$.

    HO 12

15. Suppose that for testing $H_0 : \mu = 5$ versus $H_a : \mu > 5$ of a symmetric population that has heavy tails, 16 observations are drawn at random, and we reject $H_0$ if $4(\bar{X} - 5)/S > t_{0.05,15}$. Which of the following statements is most likely to hold?

    (a) The power of the test is most likely to decrease as the sample size increases.
    (b) The sampling distribution of the test statistic will be skewed to the right.
    (c) In this scenario, chances of commiting type-II error will be smaller than the pr(Type-II error) if the observations were drawn from a normal population.
    (d) Probability of type-I error is likely to smaller than 0.05. •

(e) Keeping the sample size fixed, the power of the test is most likely to decrease as the population standard deviation decreases.

HO 12

16. Suppose $n$ observations are drawn at random from a population with unknown mean $\mu$, variance $\sigma^2$, and $\text{corr}(X_i, X_j) = \rho \in (0, 1)$ for $i \neq j$. Which of the following statements is correct?

(a) $\sigma^2$ is underestimated by $S^2$

(b) $\text{SE}(\bar{X})$ is underestimated by $S/\sqrt{n}$•

(c) $\mu$ is underestimated by $\bar{X}$

(d) The coverage probability of the confidence interval $\bar{X} \pm t_{\alpha/2,(n-1)} S/\sqrt{n}$ for $\mu$ will be larger than $100(1-\alpha)\%$

(e) $\text{var}(\bar{X}) < \sigma^2/n$

Suppose that $Y_1, \cdots, Y_{10}$ are drawn from a population. Using sign test we reject $H_0 : \tilde{\mu} = 2$ against $\tilde{\mu} > 2$ when $S^+ > k$ with a level of significance 0.05. Answer the next two questions.

17. Find the value of $k$.

    (a) 6
    (b) 7
    (c) 8 •
    (d) 9
    (e) 10

    Under $H_0$ $S^+ \sim$ Binomial$(10, 0.5)$. Thus $k$ will be determined such that $k = \min\{r : \text{pr}(S^+ > r|H_0) \leq 0.05\}$. Check $\text{pr}(S^+ > 8) = 0.0107$ and $\text{pr}(S^+ > 7) = 0.0546$. Hence answer is $k = 8$.

18. For the time being assume that $k = 8$ (this may not be the correct value of $k$). What is the power of the test when $\tilde{\mu} = 2.2$? Assume that $\text{pr}(Y > 2.2) = 0.5$ and $\text{pr}(Y > 2) = 0.56$.

    (a) 0.1111
    (b) 0.0214
    (c) 0.0107
    (d) 0.0268 •
    (e) 0.8888

    Note that here $k$ is given to you. power= $\text{pr}(\text{Reject } H_0|H_a \text{ holds}) = \text{pr}(S^+ > 8|\tilde{\mu} = 2.2)$. Observe that under the $H_a$ $\text{pr}(Y > 2) \neq 0.5$ rather it is $\text{pr}(Y > 2) = 0.56$. Therefore, when $\tilde{\mu} = 2.2$, $S^+ \sim$ Binomial$(10, 0.56)$, and consequently $\text{pr}(S^+ > 8|\tilde{\mu} = 2.2) = 0.0268$.

19. Suppose that $X_1, \cdots, X_n$ are iid observations from a Normal$(1, \sigma^2)$ distribution. Which of the following statistics is a pivot? (i.e., its distribution is free from $\sigma^2$).

    (a) $(X_1 - X_2)/(X_3 - X_4)$ •
    (b) $\sum_{i=1}^{n}(X_i - \bar{X})^2$
    (c) $\bar{X}$
    (d) $\bar{X}/\sum_{i=1}^{n} X_i^2$
    (e) $(X_1 - X_2)^2$

    $T_1 = (X_1 - X_2) \sim N(0, 2\sigma^2)$ and $T_2 = (X_3 - X_4) \sim N(0, 2\sigma^2)$, and they are independent. Also, $T_1/\sqrt{2}\sigma \sim N(0, 1)$ and $T_2/\sqrt{2}\sigma \sim N(0, 1)$ Note that $T_1/T_2 = (T_1/\sqrt{2}\sigma)/(T_2/\sqrt{2}\sigma)$. Thus it is the ratio of two normal$(0, 1)$ random variables, hence it is free from $\sigma$. Pivot idea is coming from CI handout.

    The following table summaries the outcome of a clinical trial for comparing two drug therapies for leukemia: P and PV. Twenty-one patients were assigned to drug P and forty-two patients to drug PV. Answer the next two questions.

| Drug | Success | Failure | Total |
|---|---|---|---|
| PV | 38 | 4 | 42 |
| P | 14 | 7 | 21 |
| Total | 52 | 11 | 63 |

20. The appropriate way of testing the equality of the success probabilities due to two drugs is via

    (a) Two-sample unequal variance t-test
    (b) Pooled t-test
    (c) Fisher's exact test •
    (d) Paired t-test
    (e) Pearson Chi-square test of independence

    HW problem.

21. Estimate the odds of success due to drug PV.

(a) 0.2105
(b) 4.75
(c) 2.009
(d) 0.105
(e) 9.5•

odds of success due to PV= pr(success due to PV)/$\{1 - $pr(success due to PV)$\}$ = $(38/42)/(4/42) = 38/4 = 9.5$.

A study was conducted to investigate whether there is a relationship between tonsil size and carriers of a particular bacterium, *Streptococcus pyrogenes*. The following table contains the results from 1398 children. Answer the next three questions.

| Tonsil Size | Carrier Status: Carrier | Carrier Status: Noncarrier | Row Total |
|---|---|---|---|
| Normal | 19 | 497 | 516 |
| Large | 29 | 560 | 589 |
| Very Large | 24 | 269 | 293 |
| Column Total | 72 | 1326 | 1398 |

22. What is the expected frequency for the cell 'Carrier and Normal' under the independence of carrier status and tonsil size?

(a) 516
(b) 277.91
(c) 30.33
(d) 72
(e) 26.58•

Problem from your HO.

23. What is the degrees of freedom of the Chi-square test for independence of the two factors?

(a) 6
(b) 5
(c) 3
(d) 2 •
(e) 1

df= $(3-1) \times (2-1) = 2$.

24. An alternative method of testing independence of the two factors based on the above data is

(a) Paired t-test
(b) Two-sample unequal variance t-test
(c) McNemars Test
(d) Cochran-Mantel-Haenszel Test
(e) None of the above •

See HO 13.

25. Suppose that $X_1, \cdots, X_m$ are iid from a Normal$(\mu_1, \sigma_1^2)$ distribution and $Y_1, \cdots, Y_n$ are iid from a Normal$(\mu_2, \sigma_2^2)$, and the two samples are independent. An appropriate $100(1-\alpha)\%$ confidence interval for $\rho = \sigma_2^2/\sigma_1^2$ is

(a) $\left[\frac{(n-1)S_2^2}{(m-1)S_1^2} Z_{1-\alpha/2}, \frac{(n-1)S_2^2}{(m-1)S_1^2} Z_{\alpha/2}\right]$

(b) $\left[\frac{(n-1)S_2^2}{(m-1)S_1^2\chi^2_{\alpha/2,n-1}}, \frac{(n-1)S_2^2}{(m-1)S_1^2\chi^2_{1-\alpha/2,n-1}}\right]$

(c) $\left[\frac{S_2^2}{S_1^2\chi^2_{\alpha/2,n-1}}, \frac{S_2^2}{S_1^2\chi^2_{1-\alpha/2,n-1}}\right]$

(d) $\left[\frac{S_2^2}{S_1^2\chi^2_{\alpha/2,m-1}}, \frac{S_2^2}{S_1^2\chi^2_{1-\alpha/2,m-1}}\right]$

(e) $\left[\frac{S_2^2}{S_1^2}F_{1-\alpha/2,m-1,n-1}, \frac{S_2^2}{S_1^2}F_{\alpha/2,m-1,n-1}\right]$ •

Look at HO material regarding testing of variance. Note that

$$\frac{(S_1^2/\sigma_1^2)}{S_2^2/\sigma_2^2} \sim F_{(m-1),(n-1)}$$

So,

$$\begin{aligned} & \mathrm{pr}\left\{F_{1-\alpha/2,m-1,n-1} \leq \frac{(S_1^2/\sigma_1^2)}{S_2^2/\sigma_2^2} \leq F_{1-\alpha/2,m-1,n-1}\right\} = 1-\alpha \\ \Rightarrow \quad & \mathrm{pr}\left\{F_{1-\alpha/2,m-1,n-1}\frac{S_2^2}{S_1^2} \leq \frac{\sigma_2^2}{\sigma_1^2} \leq \frac{S_2^2}{S_1^2}F_{1-\alpha/2,m-1,n-1}\right\} = 1-\alpha. \end{aligned}$$