MATH 4720 TEST 2 Mar. 13, 2017 NAME key

Please show all work for partial credit. Point totals are shown to the left of each problem.

1. An instructor hypothesizes that the standard deviation of the final exam grades in her statistics class is larger for the male students than it is for the female students. The data from the final exam for the last semester are shown. Is there enough evidence to support her claim, using $\alpha = 0.01$?

Males	Females
$n_1 = 16$	$n_2 = 16$
$\overline{x}_1 = 74$	$\overline{x}_2 = 71$
$s_1 = 4.2$	$s_2 = 2.3$

1. $H_0: \sigma_1 = \sigma_2$

   $H_a: \sigma_1 > \sigma_2$

2. $\alpha = .01$

3. $F = \frac{4.2^2}{2.3^2} = 3.33$

$F_{15,15,.01} = 3.52$

p-value = .0128

don't reject H_0

Cannot conclude Males have higher σ at $\alpha = .01$

2. Let $Y \sim GAM(\alpha = 2, \beta)$.

   a) Find the MME for β.

$E(Y) = 2\beta$

$2\beta \overset{SET}{=} \overline{Y}$

$\Rightarrow \boxed{\hat{\beta} = \overline{Y}/2}$

   b) Is the estimator you found in (a) consistent for β?

$E(\hat{\beta}) = \frac{1}{2}E(\overline{Y}) = \frac{1}{2}(\alpha\beta) = \frac{1}{2}(2)\cdot\beta = \beta$ unbiased

$V(\hat{\beta}) = \frac{1}{4}V(\overline{Y}) = \frac{1}{4}\frac{(\alpha\beta^2)}{n} = \frac{\beta^2}{2n}$

$\lim_{n\to\infty} V(\hat{\beta}) = \lim_{n\to\infty} \frac{\beta^2}{2n} = 0 \Rightarrow \boxed{\hat{\beta} \xrightarrow{P} \beta}$

3. Let $Y \sim GEO(p)$ and $p(y) = (1-p)^{y-1}p$ for $y = 1, 2, 3, \ldots$
Find the MLE for p.

$$L(p) = \prod_{i=1}^{n} (1-p)^{y_i - 1} \cdot p = p^n (1-p)^{\Sigma y_i - n}$$

$$\ln L = n \cdot \ln p + (\Sigma y_i - n) \ln(1-p)$$

$$\frac{\partial \ln L}{\partial p} = \frac{n}{p} + \frac{(\Sigma y_i - n)(-1)}{1-p} \overset{\text{SET}}{=} 0$$

$$\frac{n}{\hat{p}} = \frac{\Sigma(y_i) - n}{1 - \hat{p}} \Rightarrow n - n\hat{p} = \hat{p}(\Sigma y_i - n)$$

$$n = \hat{p}(\Sigma y_i - n + n)$$

$$\Rightarrow \hat{p} = \frac{n}{\Sigma y_i} = \boxed{\frac{1}{\bar{Y}}}$$

4. We are interested in testing $H_0: p = .6$ vs. $H_a: p \neq .6$. A random sample of size 40 is taken and $Y = no.\ of\ successes$. The rejection region is $|Y - 24| \geq 4$.
   a) Compute α. $Y \sim Bin(n = 40, p = .6)$

$$\alpha = P(Y \geq 28 \mid p = .6) + P(Y \leq 20 \mid p = .6)$$

$$= .1285 + .1298$$

$$= .2583$$

   b) Compute β if $p = .8$

$$Y \sim Bin(n = 40, p = .80)$$

$$P(21 \leq Y \leq 27) = .04324 - .0000217$$

$$= .0432$$

5. A check cashing service found that approximately 5% of all checks submitted to the service were bad. After instituting a check-verification system to reduce its losses, the service found that only 45 checks were bad in a random sample of 1124 that were cashed.
   a) Does sufficient evidence exist to affirm that the check-verification system reduce the proportion of bad checks? Use $\alpha = 0.01$.

$-z_{.01} = -2.326$

$H_0: p = .05$

$H_a: p < .05$

$\hat{p} = \frac{45}{1124} = .04$

$z = \frac{.04 - .05}{\sqrt{\frac{(.05)(.95)}{1124}}} = -1.54$

$\Rightarrow$ don't reject H_0

not enough evidence at $\alpha = .01$ to affirm the new check-verification system reduced the proportion of bad checks.

   b) What attained significance is associated with the test?

p-value $= P(Z < -1.54) = .0617$

6. Let $Y \sim BIN(1, p)$. Find the MVUE for $Var(Y) = p(1-p)$.
   - Find a sufficient statistic for p
   - Find a function of the sufficient statistic that is unbiased for $Var(Y) = p(1-p)$.

$L(p) = \prod_{i=1}^{n} p^{y_i}(1-p)^{1-y_i}, \quad y_i = 0, 1$

$= \underbrace{p^{\Sigma y_i}(1-p)^{n-\Sigma y_i}}_{g(p;\, \Sigma y_i)} \cdot \underbrace{1}_{h(y_1, \dots, y_n)}$

$\Rightarrow u = \Sigma Y_i$ is suff. for p

$E(u) = nE(Y) = np$

$\Rightarrow E(\bar{Y}) = p$

Try $E(\bar{Y}(1-\bar{Y}))$

$= E(\bar{Y} - \bar{Y}^2) = E(\bar{Y}) - E(\bar{Y}^2)$

$= \mu - (\frac{\sigma^2}{n} + \mu^2)$

$= p - [\frac{p(1-p)}{n} + p^2]$

$= p - p^2 - \frac{p(1-p)}{n}$

$= p(1-p) - \frac{1}{n}p(1-p)$

$= \frac{n-1}{n} p(1-p)$ $\rightarrow$ need to adjust

$\Rightarrow \boxed{W = \frac{n}{n-1}\bar{Y}(1-\bar{Y})}$ is MVUE for $p(1-p)$

- $E(W) = p(1-p)$
- W is a fct of a suff. stat.

7. A lumber company is interested in seeing if the number of board feet per tree has decreased since moving to a new location of timber. In the past, the company has had an average of 93 board feet per tree. The company would like to determine if average board feet per tree is different since changing locations. A random sample of 25 trees yields an average of 89 board feet per tree with a standard deviation of 20 board feet per tree. Assuming that board feet per tree are normally distributed, test the educator's claim at α = 0.10 **using a confidence interval only!** State your conclusion.

$H_0: \mu = 93$

$H_a: \mu \neq 93$

$C[82.156 < \mu < 95.6844] = .90$

$\Rightarrow$ can't conclude $\mu \neq 93$ at new location since μ is in CI

90% CI for μ:

$89 \pm 1.711 \frac{(20)}{\sqrt{25}} = 89 \pm 6.844$

$\boxed{(82.156, 95.6844)}$

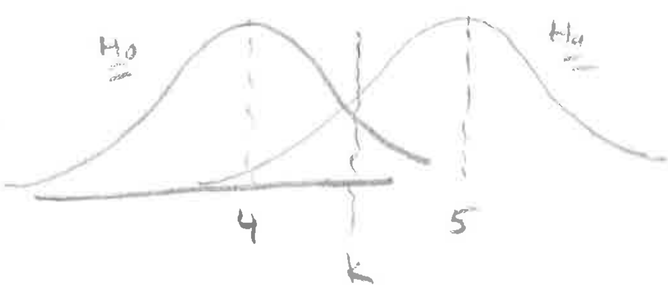
or $H_0: \mu = 93$ $H_a: \mu < 93$

90% upper CI for μ:

$89 + 1.318 \frac{20}{\sqrt{25}} = 94.272$

$C[\mu < 94.272] = .90$

8. Lowell Heiny would like to test if the average number of insect fragments in a 10 ounce sample of Heinz ketchup is more than 4. How large a sample is necessary to run a test with α = .02 and β = .10? Assume σ = 2.



H_0: $k = 4 + 2.05 \frac{2}{\sqrt{n}}$

H_a: $k = 5 - 1.282 \frac{2}{\sqrt{n}}$

$4 + 2.05 \frac{2}{\sqrt{n}} = 5 - 1.282 \frac{2}{\sqrt{n}}$

$\frac{6.664}{\sqrt{n}} = 1$ $\Rightarrow n = 6.664^2$

$= 44.4 \rightarrow 45$

9. Let $Y_1 \dots Y_{20}$ be a random sample from $Y \sim N(\mu, 100)$.
Find a uniformly most powerful test of size $\alpha = 0.05$ for H_0: $\mu = 75$ $vs.$ H_a: $\mu > 75$

$$L(\mu) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{1}{2\sigma^2}(y_i-\mu)^2} = \frac{1}{(2\pi)^{n/2}\sigma^n} \cdot e^{-\frac{1}{2\sigma^2}\sum(y_i-75)^2}$$

$$\Rightarrow \lambda = \frac{L(75)}{L(\mu_a)} = \frac{\frac{1}{(2\pi)^{n/2}\sigma^n} e^{-\frac{1}{2\sigma^2}\sum(y_i-75)^2}}{\frac{1}{(2\pi)^{n/2}\sigma^n} e^{-\frac{1}{2\sigma^2}\sum(y_i-\mu_a)^2}} = e^{-\frac{1}{2\sigma^2}\left[\sum(y_i-75)^2 + \sum(y_i-\mu_a)^2\right]} \le k$$

$$-\frac{1}{2\sigma^2}\left[\sum(y_i-75)^2 + \sum(y_i-\mu_a)^2\right] \le \ln k$$

$$\sum(y_i-75)^2 - \sum(y_i-\mu_a)^2 \ge -2\sigma^2 \ln k$$

$$\sum y_i^2 - 2(75)\sum y_i + n(75)^2 - \sum y_i^2 + 2\mu_a \sum y_i - n\mu_a^2 \ge -2\sigma^2 \cdot \ln k$$

$$2(\mu_a - 75)\sum y_i \ge -2\sigma^2 \cdot \ln k - n(75)^2 + n\mu_a^2$$

$$\sum y_i \ge \frac{-2\sigma^2 \cdot \ln k - n\left[75^2 - \mu_a^2\right]}{2(\mu_a - 75)} \quad \text{or} \quad \bar{y} \ge \underbrace{\frac{-2\sigma^2 \ln k - n\left[75^2 - \mu_a^2\right]}{2n(\mu_a - 75)}}_{c}$$

or $\boxed{\bar{y} \ge c'}$

Now, find c so that $P(\bar{y} \ge c) = .05$ if $\underline{\mu = 75}$

$$c = 75 + 1.645 \frac{(10)}{\sqrt{20}} = 78.678$$

i.e. $\boxed{\text{Reject } H_0 \text{ if } \bar{y} \ge 78.678}$