AMS 315, Spring 2021, March 4, 2021, Solution to Midterm Examination 1

AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA

1. An individual has one of three genotypes called A, B, and C, respectively, for a gene associated with disease X. The probability that an individual has genotype A is 0.5; the probability that an individual has genotype B is 0.4; and the probability that an individual has genotype C is 0.1. The probability that an individual with the A genotype is affected with disease X is 0.01. The probability that an individual with the B genotype is affected with disease X is 0.5. The probability that an individual with the C genotype is affected with disease X is 0.9.
   a. What is the probability that an individual is affected with disease X? This part of the problem is worth 10 points.
   b. Given that an individual has disease X, what is the probability that the individual is genotype B? This part is worth 40 points.

a. $P(\text{AFFECTED}) = P(\text{AFF}|A)\,P(A) + P(\text{AFF}|B)\,P(B) + P(\text{AFF}|C)\,P(C)$

$= (.01)(.5) + (.5)(.4) + (.9)(.1) =$

$= 0.005 + 0.20 + 0.09 = \boxed{0.295 = P(\text{AFFECTED})}$

b. $P(B|\text{AFFECTED}) = \dfrac{P(B \cap \text{AFFECTED})}{P(\text{AFFECTED})}$

$= \dfrac{P(\text{AFFECTED}|B)\,P(B)}{P(\text{AFFECTED})} = \dfrac{(0.5)(0.4)}{0.295} = 0.678$

$\boxed{P(B|\text{AFFECTED}) = 0.678}$

−2 COMPUTATIONAL ERROR (MINOR)

ACCEPT ANSWERS THAT ROUND OFF TO 3 DIGITS CORRECTLY.

A. CORRECT OR NOT 10 PTS

B. CORRECT OR NOT 40 PTS.

BBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBB

1. An individual has one of three genotypes called *A*, *B*, and *C*, respectively, for a gene associated with disease *X*. The probability that an individual has genotype *A* is 0.6; the probability that an individual has genotype *B* is 0.3; and the probability that an individual has genotype *C* is 0.1. The probability that an individual with the *A* genotype is affected with disease *X* is 0.1. The probability that an individual with the *B* genotype is affected with disease *X* is 0.7. The probability that an individual with the *C* genotype is affected with disease *X* is 0.9.
   a. What is the probability that an individual is affected with disease *X*? This part of the problem is worth 10 points.
   b. Given that an individual has disease *X*, what is the probability that the individual is genotype *C*? This part is worth 40 points.

1A. $P(\text{AFFECTED}) = P(\text{AFF}|A)\,P(A) + P(\text{AFF}|B)\,P(B) + P(\text{AFF}|C)\,P(C)$

$= (0.1)(0.6) + (0.7)(0.3) + (0.9)(0.1)$

$= 0.06 + .21 + .09 = 0.36.$

$\boxed{P(\text{AFFECTED}) = 0.36.}$

1B. $P(C|\text{AFFECTED}) = \dfrac{P(C \cap \text{AFFECTED})}{P(\text{AFFECTED})}$

$= \dfrac{P(\text{AFFECTED}|C)\,P(C)}{P(\text{AFFECTED})}$

$= \dfrac{(0.9)(0.1)}{0.36} = 0.25$

$\boxed{P(C|\text{AFFECTED}) = 0.25}$

−2 MINOR COMPUTATIONAL ERROR

ACCEPT ANSWERS THAT ROUND OFF TO 3 DIGITS CORRECTLY.

A CORRECT OR NOT 10 PTS

B CORRECT OR NOT 40 PTS

CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC

2. A research team took a random sample of 5 observations from the random variable Y, which had a normal distribution $N(\mu, \sigma^2)$. They observed $\bar{y}_5 = 44.2$, where $\bar{y}_5$ is the average of the 5 sampled observations and $s^2 = 138.9$ is the observed value of the unbiased estimate of σ^2, based on the sample values (i.e., the divisor in the variance was $n-1$). Test the null hypothesis $H_0: E(Y) = 60$ against $H_1: E(Y) \neq 60$. Use levels of significance 0.10, 0.05, and 0.01. This problem is worth 40 points.

$$t_4 = \frac{\bar{y}_5 - 60}{\sqrt{s^2/5}} = \frac{44.2 - 60}{\sqrt{138.9/5}} = \frac{-15.8}{5.271} = -3.00$$

α	$t_{\alpha,4}$	z_α
.10	R 2.132	1.645
.05	R 2.776	1.960
.01	A 4.604	2.576

REJECT $H_0: E(Y) = 60$ VS $H_1: E(Y) \neq 60$ AT $\alpha = .10$ AND $\alpha = .05$. ACCEPT AT $\alpha = .01$.

-30 NO DIVISION BY 5 IN STANDARD ERROR

-30 NO DECISION OR INCONSISTENT DECISION.

-25 USE Z VALUES RATHER THAN t.

-20 ONE SIDED TEST.

-15 DF ERROR

DDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDD

2. A research team took a sample of 4 observations from the random variable Y, which had a normal distribution $N(\mu, \sigma^2)$. They observed $\bar{y}_4 = 54.2$, where $\bar{y}_4$ is the average of the 4 sampled observations and $s^2 = 229.1$ is the observed value of the unbiased estimate of σ^2, based on the sample values (i.e., the divisor in the variance was $n-1$). Find the 99% confidence interval for $E(Y)$. This problem is worth 40 points.

99% CI FOR E(Y) IS

$$\bar{y}_4 \pm t_{2.576,3}\sqrt{\frac{s^2}{4}}$$

$$= 54.2 \pm 5.841\sqrt{\frac{229.1}{4}}$$

$$= 54.2 \pm 5.841\,(7.568)$$

$$= 54.2 \pm 44.2$$

$$= \boxed{10.0 \text{ TO } 98.4.}$$

-30 NO DIVISION BY 4 IN STANDARD ERROR

-25 USE Z VALUE RATHER THAN t.

-15 t VALUE OTHER THAN 5.841. (DF ERROR OR 1-α ERROR)

CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC

3. A research team took a random sample of 3 observations from a normally distributed random variable Y and observed that $\bar{y}_3 = 148.6$ and $s_Y^2 = 327.4$, where $\bar{y}_3$ was the average of the 3 observations sampled from Y and s_Y^2 was the unbiased estimate of $var(Y)$ (i.e., the divisor in the variance was $n-1$). A second research team took a random sample of 2 observations from a normally distributed random variable X and observed that $\bar{x}_2 = 89.1$ and $s_X^2 = 402.8$, where $\bar{x}_2$ was the average of the two observations sampled from X and s_X^2 was the unbiased estimate of $var(X)$ (i.e., the divisor in the variance was $n-1$). Find the 99% confidence interval for $E(X) - E(Y)$. This problem is worth 50 points.

DF $= 3+2-2$, $= 3$

$$s_p^2 = \frac{2(327.4) + 1(402.8)}{3} = \frac{1057.6}{3} = 352.53.$$

$$se(\bar{x}_2 - \bar{y}_3) = \sqrt{s_p^2\left(\tfrac{1}{3} + \tfrac{1}{2}\right)}$$

$$= \sqrt{352.53\left(\tfrac{1}{3} + \tfrac{1}{2}\right)} = \sqrt{293.775} = 17.14.$$

99% CI FOR $E(X) - E(Y)$ IS

$$\bar{x}_2 - \bar{y}_3 \pm t_{2.576,3}(17.14)$$

$$= 89.1 - 148.6 \pm 5841(17.14)$$

$$= -59.5 \pm 100.1 = -159.6 \text{ TO } 40.6.$$

+15 CORRECT S_p^2.

-35 USE NORMAL VALUES, NOT T.

-35 NO $\left(\frac{1}{n} + \frac{1}{m}\right)$ IN STANDARD ERROR.

-10 CENTER CI ON $+59.5$ RATHER THAN -59.5

ACCEPT UNEQUAL VARIANCE CI'

-20 DF ERROR.

DDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDD

3. A research team took a random sample of 6 observations from a normally distributed random variable Y and observed that $\bar{y}_6 = 148.6$ and $s_Y^2 = 227.4$, where $\bar{y}_6$ was the average of the 6 observations sampled from Y and s_Y^2 was the unbiased estimate of $var(Y)$ (i.e., the divisor in the variance was $n-1$). A second research team took a random sample of 4 observations from a normally distributed random variable X and observed that $\bar{x}_4 = 128.0$ and $s_X^2 = 302.7$, where $\bar{x}_4$ was the average of 4 observations sampled from X and s_X^2 was the unbiased estimate of $var(X)$ (i.e., the divisor in the variance was $n-1$). Test the null hypothesis $H_0: E(X) = E(Y)$ against the alternative $H_1: E(X) \neq E(Y)$ at the 0.10, 0.05, and 0.01 levels of significance using the pooled variance t-test. This problem is worth 50 points.

$$S_p^2 = \frac{5(227.4) + 3(302.7)}{8} = \frac{2045.1}{8} = 255.64$$

ON 8 DF.

$$t_8 = \frac{148.6 - 128.0 - 0}{\sqrt{255.64\left(\frac{1}{6} + \frac{1}{4}\right)}} = \frac{20.6}{\sqrt{106.52}} = \frac{20.6}{10.32}$$

$$= 1.996$$

α	Z_α	$t_{\alpha,8}$	
.10	1.645	1.860	R
.05	1.960	2.306	A
.01	2.576	3.355	A.

$$\boxed{\text{REJECT } H_0: E(X) = E(Y) \text{ VS } H_1: E(X) \neq E(Y) \text{ AT } \alpha = .10; \text{ ACCEPT AT } \alpha = .05 \text{ AND } \alpha = .01.}$$

+15 CORRECT Sp.

-35 USE NORMAL VALUES, NOT t.

-35 NO DECISION OR INCONSISTENT DECISION.

-35 NO $\left(\frac{1}{n} + \frac{1}{m}\right)$ IN STANDARD ERROR.

ACCEPT UNEQUAL VARIANCE TEST

-20 DF ERROR

FFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFF

4. In a clinical trial, *2J* patients suffering from an illness will be randomly assigned to one of two groups so that *J* will receive an experimental treatment and *J* will receive the best available treatment. The random variable *X* is the response of a patient to the experimental medicine, and the random variable *B* is the response of a patient to the best currently available treatment. Under the null hypothesis, both *X* and *B* are normally distributed with $\sigma_X = \sigma_B = 300$. The null hypothesis to be tested is that $H_0: E(X) - E(B) = 0$ against the alternative hypothesis $H_1: E(X) - E(B) > 0$ at the 0.025 level of significance.
   a. What is the number *J* in each group that would have to be taken so that the probability of a Type II error for the test of the null hypothesis specified in the common section is 0.05 when $E(X) - E(B) = 400$ and $\sigma_X = 400$ and $\sigma_B = 300$. This part is worth 45 points.
   b. What is the total number of subjects for this clinical trial? This part is worth 5 points.

a. $\sqrt{J} \geq \dfrac{|z_{\alpha}|\sqrt{\sigma_0^2 + \sigma_0^2} + |z_{\beta}|\sqrt{\sigma_0^2 + \sigma_1^2}}{\Delta}$

$\sqrt{J} \geq \dfrac{1.960\sqrt{300^2 + 300^2} + 1.645\sqrt{300^2 + 400^2}}{400}$

$\sqrt{J} \geq \dfrac{1.960(424.26) + 1.645(500)}{400}$

$\sqrt{J} \geq \dfrac{1654.05}{400} = 4.135$

$\boxed{J \geq 17.09 \quad \text{USE } J \geq 18 \text{ SUBJECTS PER GROUP}}$

b. $\boxed{\text{TOTAL IS } 2J = 36 \text{ SUBJECTS}}$

A -10 NO 1.960 -10 NO 1.645.

-20 NO $\sqrt{2(300)^2}$

-20 NO $\sqrt{300^2 + (400)^2}$

-30 NO SQUARE OF 17.09.

B. CONSISTENT OR NOT

EEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEE

4. In a clinical trial, *2J* patients suffering from an illness will be randomly assigned to one of two groups so that *J* will receive an experimental treatment and *J* will receive the best available treatment. The random variable *X* is the response of a patient to the experimental medicine, and the random variable *B* is the response of a patient to the best currently available treatment. Under the null hypothesis, both *X* and *B* are normally distributed with $\sigma_X = \sigma_B = 400$. The null hypothesis to be tested is that $H_0: E(X) - E(B) = 0$ against the alternative hypothesis $H_1: E(X) - E(B) > 0$ at the 0.005 level of significance.
   a. What is the number *J* in each group that would have to be taken so that the probability of a Type II error for the test of the null hypothesis specified in the common section is 0.01 when $E(X) - E(B) = 200$ and $\sigma_X = 500$ and $\sigma_B = 400$. This part is worth 45 points.
   b. What is the total number of subjects for this clinical trial? This part is worth 5 points.

a. $$\sqrt{J} \geq \frac{2.576\sqrt{(400)^2+(400)^2} + 2.326\sqrt{(500)^2+(400)^2}}{200}$$

$$\sqrt{J} \geq \frac{|z_\alpha|\sqrt{\sigma_0^2+\sigma_0^2} + |z_\beta|\sqrt{\sigma_0^2+\sigma_1^2}}{\Delta}$$

$$\sqrt{J} \geq \frac{2.576(565.69) + 2.326(640.31)}{200}$$

$$\sqrt{J} \geq \frac{2946.58}{200} = 14.73, \quad J \geq 217.06$$

$\boxed{\text{CHOOSE } J \geq 218 \text{ PER GROUP}}$

b. $\boxed{\text{TOTAL NUMBER IS } 2J: 436}$

A −10 NO 2.576 −20 NO $\sqrt{2(400)^2}$

−10 NO 2.326 −20 NO $\sqrt{(500)^2+(400)^2}$

−30 DON'T SQUARE 14.73.

B. CONSISTENT OR NOT

GGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG

5. A research team took a random sample of 4 observations from a normally distributed random variable Y and observed that $\bar{y}_4 = 831.2$ and $s_Y^2 = 5{,}263.2$, where $\bar{y}_4$ was the average of the 4 observations sampled from Y and s_Y^2 was the unbiased estimate of $var(Y)$. A second research team took a random sample of 5 observations from a normally distributed random variable X and observed that $\bar{x}_5 = 248.9$ and $s_X^2 = 3154.4$, where $\bar{x}_5$ was the average of the 5 observations sampled from X and s_X^2 was the unbiased estimate of $var(X)$. Find the 95% confidence interval for $\frac{var(X)}{var(Y)}$. This problem is worth 50 points.

SINCE $\frac{VAR(X)}{VAR(Y)}$ IS THE PARAMETER TARGET,

TS HAS AN F DIST WITH 3, 4 DF.

NOW $F_{0.025,3,4} = 9.98$

AND $F_{0.025,4,3} = 15.10$ ARE BOUNDARY

PERCENTILES FOR A 95% CI.

$$\frac{s_X^2}{s_Y^2} = \frac{3{,}154.4}{5{,}263.2} = 0.599.$$

THE 95% CI FOR $\frac{VAR(X)}{VAR(Y)}$ IS

$$\frac{1}{15.10}(0.599) = 0.0397 \text{ TO}$$

$$9.98\,(0.599) = 5.98$$

-50 TWO SAMPLE T TEST.

-15 EACH DF ERROR EXCEPT REVERSAL

-20 REVERSAL OF DF'S.

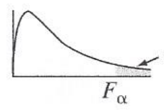

TABLE 8

Percentage points of the F distribution (df_2 between 1 and 6)

		df_1									
df_2	α	1	2	3	4	5	6	7	8	9	10
1	.25	5.83	7.50	8.20	8.58	8.82	8.98	9.10	9.19	9.26	9.32
	.10	39.86	49.50	53.59	55.83	57.24	58.20	58.91	59.44	59.86	60.19
	.05	161.4	199.5	215.7	224.6	230.2	234.0	236.8	238.9	240.5	241.9
	.025	647.8	799.5	864.2	899.6	921.8	937.1	948.2	956.7	963.3	968.6
	.01	4052.2	4999.5	5403.3	5624.6	5763.7	5859.0	5928.4	5981.0	6022.5	6055.8
2	.25	2.57	3.00	3.15	3.23	3.28	3.31	3.34	3.35	3.37	3.38
	.10	8.53	9.00	9.16	9.24	9.29	9.33	9.35	9.37	9.38	9.39
	.05	18.51	19.00	19.16	19.25	19.30	19.33	19.35	19.37	19.38	19.40
	.025	38.51	39.00	39.17	39.25	39.30	39.33	39.36	39.37	39.39	39.40
	.01	98.50	99.00	99.17	99.25	99.30	99.33	99.36	99.37	99.39	99.40
	.005	198.5	199.0	199.2	199.2	199.3	199.3	199.4	199.4	199.4	199.4
	.001	998.5	999.0	999.2	999.2	999.3	999.3	999.4	999.4	999.4	999.4
3	.25	2.02	2.28	2.36	2.39	2.41	2.42	2.43	2.44	2.44	2.44
	.10	5.54	5.46	5.39	5.34	5.31	5.28	5.27	5.25	5.24	5.23
	.05	10.13	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81	8.79
	.025	17.44	16.04	15.44	15.10	14.88	14.73	14.62	14.54	14.47	14.42
	.01	34.12	30.82	29.46	28.71	28.24	27.91	27.67	27.49	27.35	27.23
	.005	55.55	49.80	47.47	46.19	45.39	44.84	44.43	44.13	43.88	43.69
	.001	167.0	148.5	141.1	137.1	134.6	132.8	131.6	130.6	129.9	129.2
4	.25	1.81	2.00	2.05	2.06	2.07	2.08	2.08	2.08	2.08	2.08
	.10	4.54	4.32	4.19	4.11	4.05	4.01	3.98	3.95	3.94	3.92
	.05	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00	5.96
	.025	12.22	10.65	9.98	9.60	9.36	9.20	9.07	8.98	8.90	8.84
	.01	21.20	18.00	16.69	15.98	15.52	15.21	14.98	14.80	14.66	14.55
	.005	31.33	26.28	24.26	23.15	22.46	21.97	21.62	21.35	21.14	20.97
	.001	74.14	61.25	56.18	53.44	51.71	50.53	49.66	49.00	48.47	48.05
5	.25	1.69	1.85	1.88	1.89	1.89	1.89	1.89	1.89	1.89	1.89
	.10	4.06	3.78	3.62	3.52	3.45	3.40	3.37	3.34	3.32	3.30
	.05	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77	4.74
	.025	10.01	8.43	7.76	7.39	7.15	6.98	6.85	6.76	6.68	6.62
	.01	16.26	13.27	12.06	11.39	10.97	10.67	10.46	10.29	10.16	10.05
	.005	22.78	18.31	16.53	15.56	14.94	14.51	14.20	13.96	13.77	13.62
	.001	47.18	37.12	33.20	31.09	29.75	28.83	28.16	27.65	27.24	26.92
6	.25	1.62	1.76	1.78	1.79	1.79	1.78	1.78	1.78	1.77	1.77
	.10	3.78	3.46	3.29	3.18	3.11	3.05	3.01	2.98	2.96	2.94
	.05	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10	4.06
	.025	8.81	7.26	6.60	6.23	5.99	5.82	5.70	5.60	5.52	5.46
	.01	13.75	10.92	9.78	9.15	8.75	8.47	8.26	8.10	7.98	7.87
	.005	18.63	14.54	12.92	12.03	11.46	11.07	10.79	10.57	10.39	10.25
	.001	35.51	27.00	23.70	21.92	20.80	20.03	19.46	19.03	18.69	18.41

Source: Computed by M. Longnecker using the R function qf($1 - \alpha$, df_1, df_2).
Additional values can be obtained using the same R function.

HHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH

5. A research team took a random sample of 7 observations from a normally distributed random variable Y and observed that $\bar{y}_7 = 2{,}431.2$ and $s_Y^2 = 3{,}263.2$, where $\bar{y}_7$ was the average of the 7 observations sampled from Y and s_Y^2 was the unbiased estimate of $var(Y)$. A second research team took a random sample of 6 observations from a normally distributed random variable X and observed that $\bar{x}_6 = 248.9$ and $s_X^2 = 4{,}315.4$, where $\bar{x}_6$ was the average of the 6 observations sampled from X and s_X^2 was the unbiased estimate of $var(X)$. Find the 99% confidence interval for $\frac{var(X)}{var(Y)}$. This problem is worth 50 points.

SINCE $\frac{VAR(X)}{VAR(Y)}$ IS TARGET OF CI.

TS $\sim F(6,5)$

$F_{.005,6,5} = 14.51$

$F_{.005,5,6} = 11.46$.

$$\frac{s_X^2}{s_Y^2} = \frac{4315.4}{3263.2} = 1.32$$

99% CI FOR $\frac{VAR(X)}{VAR(Y)}$ IS

$\frac{1}{11.46}(1.32)$ TO $14.51(1.32)$;

THIS IS. 0.115 TO 19.2. FOR THE 99% CI FOR VAR(X)/VAR(Y)

-50 TWO SAMPLE TEST.

-15 EACH DF ERROR EXCEPT REVERSAL.

-20 REVERSAL OF DF'S.

IIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIII

6. The random variable Z has expected value $E(Z) = 0$, $var(Z) = 1$, $E(Z^3) = 0$, and $E(Z^4) = 3$. Let Z_1 and Z_2 be a random sample of size 2 from Z. Let $T = Z_1^2 + Z_2^2$.
   a. Find $E(T)$. This part is worth 10 points.
   b. Find $var(T)$. This part is worth 50 points.

End of the Examination

a. $E(Z^2) = VAR(Z) + [E(Z)]^2 = 1+0 = 1.$

$E(T) = E(Z_1^2 + Z_2^2) = E(Z_1^2) + E(Z_2^2) = 1+1 = 2.$

$\boxed{E(T)=2}$ 10 PTS CORRECT OR NOT

b. $VAR(Z^2) = E\{[Z^2 - E(Z^2)]^2\}.$

$= E\{Z^4 - 2[E(Z^2)]Z^2 + [E(Z^2)]^2\}$

$= E\{Z^4\} - (E(Z^2))^2 = E(Z^4) - [E(Z^2)]^2$

$= 3 - 1^2 = 2.$ +10 POINTS EACH CORRECT STEP

$VAR(T) = VAR(Z_1^2 + Z_2^2)$

$= VAR(Z_1^2) + VAR(Z_2^2) + 2COV(Z_1^2, Z_2^2).$

$= 2 + 2 - 2 \cdot 0 = 4$

SINCE $COV(Z_1^2, Z_2^2) = 0$

$\boxed{VAR(T) = 4}$

ALSO $T = Z_1^2 + Z_2^2$ IS A CHI-SQUARE DIST WITH 2 DF. HENCE $E(T) = $ DF OF $T = 2$.

AND $VAR(T) = 2 \times (\text{DF OF } T) = 2 \times 2 = 4.$