BST 140.652 Midterm Exam
Notes:

- You may not use a calculator for this exam.
- You may use your single formula sheet.
- Please be neat and write legibly. Use the back of the pages if necessary.
- Good luck!

____

printed name

Notes:
There are 7 questions. You may omit one question. (Hence, complete 6 and omit one of your choosing.) Please mark the question that you are omitting clearly.

Also, clearly denote your answers and differentiate them from your scratch work.

All information given in an answer will be graded.

1. Researchers are studying the relative concentration of blood lead for factory workers. They took the natural logarithm of ratio of blood lead concentration for 8 factory workers and 8 control subjects. The measurements resulted in a mean log concentration of 6 (log parts per volume) for the factory workers and 4 for the control subjects. The sample variance in the factory workers was 3 while it was 5 in the control group.

   (a) Create and interpret a 95% confidence interval for the difference in the population mean of log blood lead concentration between factory workers and controls. (Assume a common variance.)

   (b) State the assumptions used to create this interval.

Note equal s.s.

$$S_p = \sqrt{\frac{3}{2} + \frac{5}{2}} = 2 \qquad t_{14,.975} = 2.14$$

C.I.

$$6 - 4 \pm 2.14\left(2\left(\frac{1}{8} + \frac{1}{8}\right)^{1/2}\right)$$

$$2 \pm 2.14 = [-.14, 4.14]$$

We estimate that the pop. diff in mean log blood concentration is betwee [-.14, 4.14] with 95% confidence. As 0 is in the interval this suggests a non significant two sided hypothesis test of equivalent means between the two groups (at the $\alpha = .05$ level for a two-sided test)

Assumption -
dat are iid normal within groups with a common variance across groups.

2. Refer to the previous problem. Let $X_{L,i}$ be the *natural-scale* blood lead measurement for subject i from the population of lead workers and $X_{C,i}$ be *natural-scale* blood lead measurement for subject i from the population of control subjects. Let $Y_{L,i} = \log(X_{L,i})$ and $Y_{C,i} = \log(X_{C,i})$. Assume that you have n subjects from each group. Let $\mu_L = E[X_{L,i}]$, $\lambda_L = E[Y_{L,i}]$, $\mu_C = E[X_{C,i}]$ and $\lambda_C = E[Y_{C,i}]$. Let $\bar{X}_L$ be the average of the $X_{L,i}$, $\bar{Y}_L$ be the average of the $Y_{L,i}$, let $\bar{X}_C$ be the average of the $X_{C,i}$ and $\bar{Y}_C$ be the average of the $Y_{C,i}$. Answer the following:

   (a) Write out the ratio of the sample geometric means between lead workers and controls using the notation given.

   (b) Using the notation given, what would the ratio of geometric means converge to as n gets large?

   (c) Suppose you were to exponentiate the endpoints of the interval from the previous problem, using the notation given, what are you estimating?

   (d) Using the notation given, as n gets very large what would $\log\left\{\frac{\bar{X}_L}{\bar{X}_C}\right\}$ become more like?

(a) $\exp\{\bar{y}_L - \bar{y}_C\}$

(b) $\exp\{\lambda_L - \lambda_C\}$ = ratio of pop. geometric means

(c) Note data was logged so it would estimate $\frac{\exp\{\lambda_L\}}{\exp\{\lambda_C\}}$ where e^{λ_L} is the pop geometric mean blood lead concentration among lead workers, and e^{λ_C} is to the controls.

(d) $\log\left\{\frac{\bar{\mu}_L}{\bar{\mu}_C}\right\}$, the log ratio of the means.

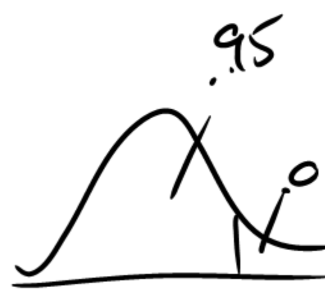
3. Refer to problem 1. Test the hypothesis that the population mean of the log of blood lead concentration for factory workers is higher than that of controls. State your hypotheses using the notation from problem 2 and write out a conclusion within the context of the problem.

$H_0: \lambda_L = \lambda_c \qquad H_a: \lambda_L \geq \lambda_c$

using notation from previous page

$$T\text{ stat} = \frac{\bar{y}_L - \bar{y}_c}{s_p\sqrt{\frac{1}{n_L} + \frac{1}{n_c}}} = \frac{6-4}{2\sqrt{\frac{1}{8}+\frac{1}{8}}} = 2$$

Compare 2 to t-stat



$t_{14,.95}$ 1.76

We reject the null hypothesis and conclude that there is sufficient evidence to suggest a difference in the mean log blood lead concentration between lead workers & controls ($\alpha = .05$ for one sided test)

4. Intelligent quotients in the general population have a mean of 100. Researchers are looking at a sub-population of interest and hypothesize that their mean IQ is higher. A sample of 100 subjects yielded an average IQ of 101.8 and a sample variance of 144. Test the relevant hypothesis defining any notation that you use. Report a P-value and report your results in the context of the problem.

$H_0: \mu = 100$ $\quad H_a: \mu > 100$

μ = mean IQ in the new population

$$T.S. = \frac{101.8 - 100}{\sqrt{\frac{144}{100}}} = \frac{1.8}{1.2} = 1.5$$

Compare to $Z_{.95} = 1.645$

We fail to reject and conclude that there is insufficient support for the hypothesis that the mean IQ in this sub pop. is larger than 100 (one sided test $\alpha = .05$)

Solving this as written

5. A new diet pill claims to cause a decrease in weight. In a one year program, the manufacturers hypothesize that their sample of 225 subjects (all with the same starting weights) will lose a mean of 5 pounds with a standard deviation of weight change (baseline - followup) of 25 pounds. Assuming a 5% type I error rate for the relevant one sided test, what is the probability of rejecting the relevant null hypothesis of a non-effective pill if their claims are true?

Relevant null $H_0: \mu = 0$

$H_a: \mu > 0$ $\quad Z_{.95} \approx 1.65$

Where μ = mean weight change (BL - FU)

Consider specific $H_a: \mu = 5$

$$P\left(\frac{\bar{X}}{\sigma/\sqrt{n}} \geq 1.65 \mid H_a \right)$$

$$\frac{5}{\sigma/\sqrt{n}} = \frac{5}{25/15} = 3$$

$$= P\left(\frac{\bar{X} - 5}{\sigma/\sqrt{n}} \geq 1.65 - \frac{5}{\sigma/\sqrt{n}} \mid H_a \right)$$

$$= P(Z \geq 1.65 - 3) = P(Z \geq -1.35)$$

$\approx 91\%$

6. You collect an iid sample from a population and obtain the data (in ascending order):

$$1,\ \cancel{2},\ 4,$$

(a) List out all of the equally likely bootstrap resamples from this data and calculate the median of each.

(b) Calculate the bootstrapped distribution of the sample median using your answer from question (a).

a

resample

(1, 1) (1, 4) (4, 1) (4, 4)

1 2.5 2.5 4

Median

b

Med	1	2.5	4
P	.25	.50	.25

7. Recall that the Poisson distribution is $P(X = x) = \lambda^x e^{-\lambda}/x!$ for $x = 0, 1, 2, \ldots$ and $\lambda > 0$ is the mean, $E[X] = \lambda$. Consider writing an exponential prior on lambda $f(\lambda) = \beta e^{-\lambda\beta}$ for $\lambda > 0$, where $\beta > 0$ is a specified number. The mean of this distribution is β^{-1}. Suppose that you collect data and obtain $x = 3$.

   (a) Write out the likelihood for λ.

   (b) Write out the posterior for λ; what distribution is it? (Note, I'm not asking you to calculate the distribution function, I'm asking what the name of the distribution is.)

   (c) What is the posterior mean?

(a) $\mathcal{L}(\lambda; x=3) = \lambda^3 e^{-3}$

(b) Post $\propto$ Prior * Likelihood

$$= \beta e^{-\lambda\beta} * \lambda^3 e^{-3} \propto \lambda^3 e^{-\lambda(\beta+3)}$$

Gamma (we gave credit for a lot of different answers here)

(c) Symbolic answer

$$E[\lambda \mid X=3] = \int \lambda \text{ posteriorpdf}(\lambda)\, d\lambda$$

we gave credit for getting this far,

Below I'm ~~how~~ how you could arrive at the exact answer using only what you

Know from this class.

Poserior of $(\lambda) \propto \lambda^3 e^{-\lambda(\beta+3)}$

Hence exact posterior is $= \dfrac{\lambda^3 e^{-\lambda(\beta+3)}}{\int_0^\infty t^3 e^{-t(\beta+3)} dt}$

Hence

$$E[\lambda | X=3] = \frac{\int_0^\infty \lambda^4 e^{-\lambda(\beta+3)} \; *}{\int_0^\infty t^3 e^{-t(\beta+3)} dt}$$

set $u = \lambda^4$ $\quad dv = e^{-\lambda(\beta+3)} \Rightarrow v = \frac{-1}{(\beta+3)} e^{-\lambda(\beta+3)}$

$du = 4\lambda^3$

$\int u\,dv = uv - \int v\,du$

Hence $* = \dfrac{-\lambda^4 e^{-\lambda(\beta+3)}}{(\beta+3)} \Big|_0^\infty + \int_0^\infty \dfrac{4}{\beta+3} \lambda^3 e^{-\lambda(\beta+3)} d\lambda$

$$= \left(\frac{4}{\beta+3}\right) \int_0^\infty \lambda^3 e^{-\lambda(\beta+3)} d\lambda$$

putting * back in we see that

$$E[\lambda \mid x=3] = \frac{\left(\frac{4}{\beta+3}\right) \int_0^\infty \lambda^3 e^{-\lambda(\beta+3)} \, d\lambda}{\int_0^\infty t^3 e^{-t(\beta+3)} \, d\lambda}$$

$$= \frac{4}{\beta+3}$$

I know it was a hard question. This is basically why I allowed the retake.