INSTRUCTIONS: Solve 8 of the following problems now and write your solutions on the provided paper, clearly labeling each solution. Take this sheet with you and solve the remaining 5 problems at home. Those solutions are due on Wednesday, February 24. Write your solutions clearly and explain proofs or other arguments using English words and sentences. You may use a calculator during all portions of the exam. Books, notes, and other non-living resources may be used on the take-home portion of the exam, but you should cite any sources that are not normal class material (e.g. websites).

1. A random sample of size $n = 100$ is to be taken from a Poisson population and used to test the null hypothesis that $\lambda = 5$ against the alternative hypothesis that $\lambda = 4$. The null hypothesis is to be rejected in favor of the alternative if $\overline{x} < 4.4$. What is the probability of a type II error?

2. I have a hypothesis that 10% of male Gonzaga students are named Michael and I have decided to test my hypothesis using a random sample of 20 male Gonzaga students. I will reject my hypothesis if none of the students in my sample are named Michael. What is the probability of a type I error?

3. The random variables X and Y have joint density

$$f(x,y) = \begin{cases} \frac{2}{5}(2x+3y) & \text{for } 0 < x < 1 \text{ and } 0 < y < 1 \\ 0 & \text{elsewhere} \end{cases}$$

Calculate the regression equation of Y on X.

4. Suppose that a random sample of 20 bottles of Cynar (a bitter Italian liqueur made from artichokes) is selected and the alcohol content (proof) of each bottle is determined. Let μ be the true mean alcohol content of Cynar. Suppose that the 95% confidence interval for μ calculated from the sample data is $(31.4, 33.9)$.

a) Would a 99% confidence interval calculated from the same data be wider or narrower than the 95% confidence interval? Explain your answer.

b) Consider the following statement: We are confident that 95% of all bottles of Cynar are between 31.4 and 33.9 proof. Is this statement correct? Explain your answer.

5. Suppose that the weight of adult rats is normally distributed. A random sample of 9 rats from a certain population have a mean weight of $\overline{x} = 220$g and a variance of $s^2 = 3600$g. Calculate a 98% confidence interval for the mean weight of rats in this population.

6. In a survey of 1600 American adults, 272 were found to be seriously overweight. Calculate a 99% confidence lower bound for the proportion of American adults who are overweight.

7. In an investigation of the toxin produced by a poisonous snake, a researcher collected 25 samples of the toxin, each weighing 1 g, and then measured the amount of antitoxin needed to neutralize the toxin. The sample variance was $s^2 = .1762$ mg. Assuming that the amount of antitoxin needed to neutralize the toxin is normally distributed, calculate a 95% confidence upper bound for the variance of the amount of antitoxin needed.

8. The authors of the article "Boredom in Young adults–Gender and Cultural Comparisons" administered the Boredom Proneness Scale to 97 male and 148 female U. S. college students. A summary of the results is shown below (a higher score means a increased Proneness to Boredom). Does the accompanying data support the conclusion that male college students are more prone to boredom than female college students? Calculate a test statistic and the P-value of the test statistic.

Gender	Sample Size	Sample Mean	Sample SD
Male	97	10.40	4.83
Female	148	9.26	4.68

9. One technology proposed for the rehabilitation of municipal water pipelines uses a polyethelyne liner threaded through existing pipes. The article "Effect of Welding on a High-Density Polyethelyne Liner" reports the following data on the tensile strength (psi) of liners subjected to a certain fusion process and on untreated liners.

Process	Sample Size	Sample Mean	Sample SD
Fusion	8	3108.1	205.9
No fusion	10	2902.8	277.3

Let σ_1^2 and σ_2^2 be the true variance of the tensile strength of the fused and non-fused liners, respectively. Assuming that tensile strength of the liners is normally distributed, test $H_0 : \frac{\sigma_1^2}{\sigma_2^2} = 1$ against $H_1 : \frac{\sigma_1^2}{\sigma_2^2} < 1$ at a significance level of $\alpha = 0.05$.

10. A random sample of size $n = 64$ is used to test the hypothesis $H_0 : \mu = 20$ against the alternate hypothesis $H_1 : \mu \neq 20$. The sample mean was $\overline{x} = 18.6$ and the sample variance was $s^2 = 31.36$. Calculate the P-value of the sample.

11. Two independent random samples are taken from a single normally distributed population in order to test the hypothesis $\mu = 0$ against the alternative $\mu \neq 0$. The first sample, with size $n = 4$, has a mean of $\overline{x} = 2.353$ and a standard deviation of $s = 2$. The second sample, with size $n = 61$, has a mean of $\overline{x} = 1.024$ and a standard deviation of $s = 4$. Which result is stronger evidence in favor of the alternative hypothesis? You should be able to estimate P-values using the table of t-critical values.

12. An epidemiologist is trying to discover the cause of a certain kind of cancer. He studies a group of 10,000 people for five years, measuring 48 different "factors" involving eating habits, drinking habits, exercise, and so on. His object is to determine for each factor if there is a difference between the mean value of the factor for those who developed cancer and for those who did not (his null hypothesis is that the factors and cancer are all independent). In an effort to be cautiously conservative, he uses a significance level of $\alpha = 0.01$ in all his tests. What is the probability that at least one of the factors will be found to be associated with the cancer, even if none of them is actually connected (that is, what is the probability of making at least one type I error)?

13. In our example of hours studied for an exam (x) and scores on that exam (y) we found that $\overline{x} = 10$, $\overline{y} = 56.4$, $S_{xx} = 376$, $S_{xy} = 1305$, and $S_{yy} = 4752.4$. Estimate the number of hours a student who scored an 85 on the exam spent studying.

Test statistics	$100(1-\alpha)\%$ Confidence interval
$z = \frac{\overline{x}-\mu_0}{\frac{\sigma}{\sqrt{n}}}$	$\overline{x} \pm z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$
$z = \frac{\overline{x}-\mu_0}{\frac{s}{\sqrt{n}}}$	$\overline{x} \pm z_{\frac{\alpha}{2}} \frac{s}{\sqrt{n}}$
$t = \frac{\overline{x}-\mu_0}{\frac{s}{\sqrt{n}}}$	$\overline{x} \pm t_{\frac{\alpha}{2},n-1} \frac{s}{\sqrt{n}}$
$z = \frac{\hat{\theta}-\theta_0}{\sqrt{\frac{\theta_0(1-\theta_0)}{n}}} = \frac{x-n\theta_0}{\sqrt{n\theta_0(1-\theta_0)}}$	$\hat{\theta} \pm z_{\frac{\alpha}{2}} \sqrt{\frac{\theta_0(1-\theta_0)}{n}}$
$\chi^2 = \frac{(n-1)s^2}{\sigma_0^2}$	$\left(\frac{(n-1)s^2}{\chi^2_{\frac{\alpha}{2},n-1}}, \frac{(n-1)s^2}{\chi^2_{1-\frac{\alpha}{2},n-1}}\right)$
$z = \frac{\overline{x}_1-\overline{x}_2-\delta_0}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}}$	$\overline{x}_1-\overline{x}_2 \pm z_{\frac{\alpha}{2}} \sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}$
$z = \frac{\overline{x}_1-\overline{x}_2-\delta_0}{\sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}}}$	$\overline{x}_1-\overline{x}_2 \pm z_{\frac{\alpha}{2}} \sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}}$
$t = \frac{\overline{x}_1-\overline{x}_2-\delta_0}{s_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}$ where $s_p^2 = \frac{(n_1-1)s_1^2+(n_2-1)s_2^2}{n_1+n_2-2}$	$\overline{x}_1-\overline{x}_2 \pm t_{\frac{\alpha}{2},n_1+n_2-2} s_p \sqrt{\frac{1}{n_1}+\frac{1}{n_2}}$
$f = \frac{s_1^2}{s_2^2}$ (which has n_1-1, n_2-1 df)	

Estimators (and test statistics) for linear regression:

$$\hat{\beta} = \frac{S_{xy}}{S_{xx}}$$

$$\hat{\alpha} = \overline{y} - \hat{\beta}\overline{x}$$

$$S_{xx} = \sum_{i=1}^{n}(x_i-\overline{x})^2 = \sum_{i=1}^{n} x_i^2 - \frac{1}{n}\left(\sum_{i=1}^{n} x_i\right)^2$$

$$S_{xy} = \sum_{i=1}^{n}(x_i-\overline{x})(y_i-\overline{y}) = \sum_{i=1}^{n} x_i y_i - \frac{1}{n}\left(\sum_{i=1}^{n} x_i\right)\left(\sum_{i=1}^{n} y_i\right)$$

$$S_{yy} = \sum_{i=1}^{n}(y_i-\overline{y})^2 = \sum_{i=1}^{n} y_i^2 - \frac{1}{n}\left(\sum_{i=1}^{n} y_i\right)^2$$

In normal regression analysis we also have the estimator

$$\hat{\sigma} = \sqrt{\frac{S_{yy}-\hat{\beta}S_{xy}}{n}},$$

the test statistic

$$t = \frac{\hat{\beta}-\beta}{\hat{\sigma}}\sqrt{\frac{(n-2)S_{xx}}{n}},$$

and the $100(1-\alpha)\%$ confidence interval for β:

$$\hat{\beta} \pm t_{\frac{\alpha}{2},n-2}\hat{\sigma}\sqrt{\frac{n}{(n-2)S_{xx}}}$$