1. Suppose you have a random sample of size n from a $N(\mu, \sigma)$ PDF. The Neyman-Pearson Lemma gives the following 4 statistics (or slight variations thereof) as the best use of the data under various assumptions:

$$(a)\ \frac{\bar{X}-\mu_0}{\sigma/\sqrt{n}} \qquad (b)\ \frac{\bar{X}-\mu_0}{S/\sqrt{n}}$$
$$(c)\ \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \qquad (d)\ \frac{\bar{X}-\mu}{S/\sqrt{n}}\,.$$

Explain when and how you should use these statistics appropriately for testing and reporting results.

Hint: State the purpose and assumptions about the normal parameters.

2. Consider the Pearson father-son height data as a sample of $n = 1{,}078$ bivariate normal data points (X_i, Y_i). The summary statistics are

$$\bar{x} = 67.68710 \qquad \sum_{i=1}^{1078}(x_i-\bar{x})^2 = 8114.44391$$
$$\bar{y} = 68.68407 \qquad \sum_{i=1}^{1078}(y_i-\bar{y})^2 = 8532.58103$$
$$\sum_{i=1}^{1078}(x_i-\bar{x})(y_i-\bar{y}) = 4171.57912$$

(a) What is the sample correlation coefficient between X and Y?

(b) Genetic theory suggests that the theoretical correlation should be 0.5. Test the null hypothesis that $H_0 : \rho = 0.5$ versus $H_1 : \rho \neq 0.5$ at the $\alpha = 5\%$ level.

(c) What is the p-value?

(d) What is the 95% confidence interval for ρ?

3. Consider the Pearson χ^2 test of independence of row and column categories in an $r \times s$ table, where the observed counts are $\{o_{ij}\}$. Of course, $\sum_{i=1}^{r}\sum_{j=1}^{s} o_{ij} = n$.

(a) If you use MLE to estimate the unknown probabilities p_{ij} in order to estimate the expected values e_{ij} as $n\hat{p}_{ij}$, show that the row and column totals exactly match the observed counts; that is,

$$\sum_{i=1}^{r} e_{ij} = \sum_{i=1}^{r} o_{ij} \qquad \text{and} \qquad \sum_{j=1}^{s} e_{ij} = \sum_{j=1}^{s} o_{ij}$$

(b) Use this fact to prove the computational identity

$$\sum_{i=1}^{r}\sum_{j=1}^{s}\frac{\left(o_{ij}-e_{ij}\right)^2}{e_{ij}} = \sum_{i=1}^{r}\sum_{j=1}^{s}\frac{o_{ij}^2}{e_{ij}} - n\,.$$

4. **Designing an Experiment:** As a statistical consultant, you are working with a Rice bioengineer who has invented new materials and processes for artificial bones. She wishes to test if the new concoction is stronger than existing materials. She knows that current commercially available materials will fracture when increasing force is applied with mean force of $\mu = 4000$ newtons, with a standard deviation σ of 300 newtons.

   She hopes her new material will fracture at a higher average of 4200 (or more). She plans to manufacture n samples with her new material and measure X_i, the force at which the i^{th} bone sample fails.

   (a) You convince her to set the type I error to 5% and the power function at $\mu = 4200$ to be (at least) 85%. What assumptions would you make about the parametric form of the data density that are realistic but as simple as possible?

   Hint: Simple might mean 2 simple hypotheses: $\mu_0 = 4000$ *and* $\mu_1 = 4200$.

   (b) Assuming you advocated the model $N(\mu, 300^2)$, what is the smallest sample size that would achieve the desired power and type I error?

   Hint: You will also need to find the critical region to answer this question. You might trying looping over possible values of n.

   (c) If you were going to recommend a composite alternative hypothesis, would you advocate a one-sided or two-sided alternative? Why?

5. **Best Critical Region:** Suppose you have a random sample $\boldsymbol{X} = \{X_1, X_2, \ldots, X_n\}$ from the double-exponential PDF

$$f(x|\theta) = \frac{1}{2\theta} e^{-|x|/\theta}, \qquad -\infty < x < \infty .$$

   Find the log-likelihoods $\ell(\theta_0|\boldsymbol{X})$ and $\ell(\theta_1|\boldsymbol{X})$, compute the log-likelihood-ratio, and simplify to find **the form of the best statistic** to perform the **test of the simple hypothesis** $H_0 : \theta = \theta_0$ versus **the simple alternative hypothesis** $H_1 : \theta = \theta_1 \neq \theta_0$.

   Hints: (i) Note that the PDF has the absolute value of x *in the exponential; (ii) Your answer may depend on the relative position of* θ_1 *to* θ_0*; and (iii) You do not need to give an expression for finding the cutoff* k'.

6. For arbitrary constants a, b, c, and d, show

$$\mathrm{Cov}(a + bX, c + dY) = b \cdot d \cdot \mathrm{Cov}(X, Y) .$$

7. **ANOVA:** A small brewery decided to experiment with label designs on its summer ale. Nineteen stores of similar size were selected. The number of cases sold over a two-month period were recorded:

Design						Total
1	11	17	16	14	15	73
2	12	10	15	19	11	67
3	23	20	18	17		78
4	27	33	22	26	28	136

(a) (10 points) Test the null hypotheses that the mean number of sales at a store are the same for all four label designs, versus the alternative that some are different. Use $\alpha = 1\%$.

(b) What is the p-value?

8. Consider a test of the hull hypothesis that

$$H_0 : \mu_x = \mu_y \quad \text{versus} \quad H_1 : \mu_x \neq \mu_y ,$$

assuming normal populations with unknown parameters. This can be tested either with a two-sample T-test, or using an ANOVA with $m = 2$ populations. In fact, these 2 approaches give exactly the same results. As random variables using Equations (7.19) (cf (7.13)) and (8.41) these 2 tests take the form

$$T_{n-2} = \frac{N(0,1)}{\sqrt{\chi^2(n-2)/(n-2)}}$$
$$F_{1,n-2} = \frac{\chi^2(1)/1}{\chi^2(n-2)/(n-2)} .$$

Explain why T_{n-2}^2 gives the same test results.

9. **Space Shuttle Flight 25** The 23 data points were analyzed either including or excluding the launches where there were no O-Ring failures. We repeat those two analyses. The temperatures recorded were:

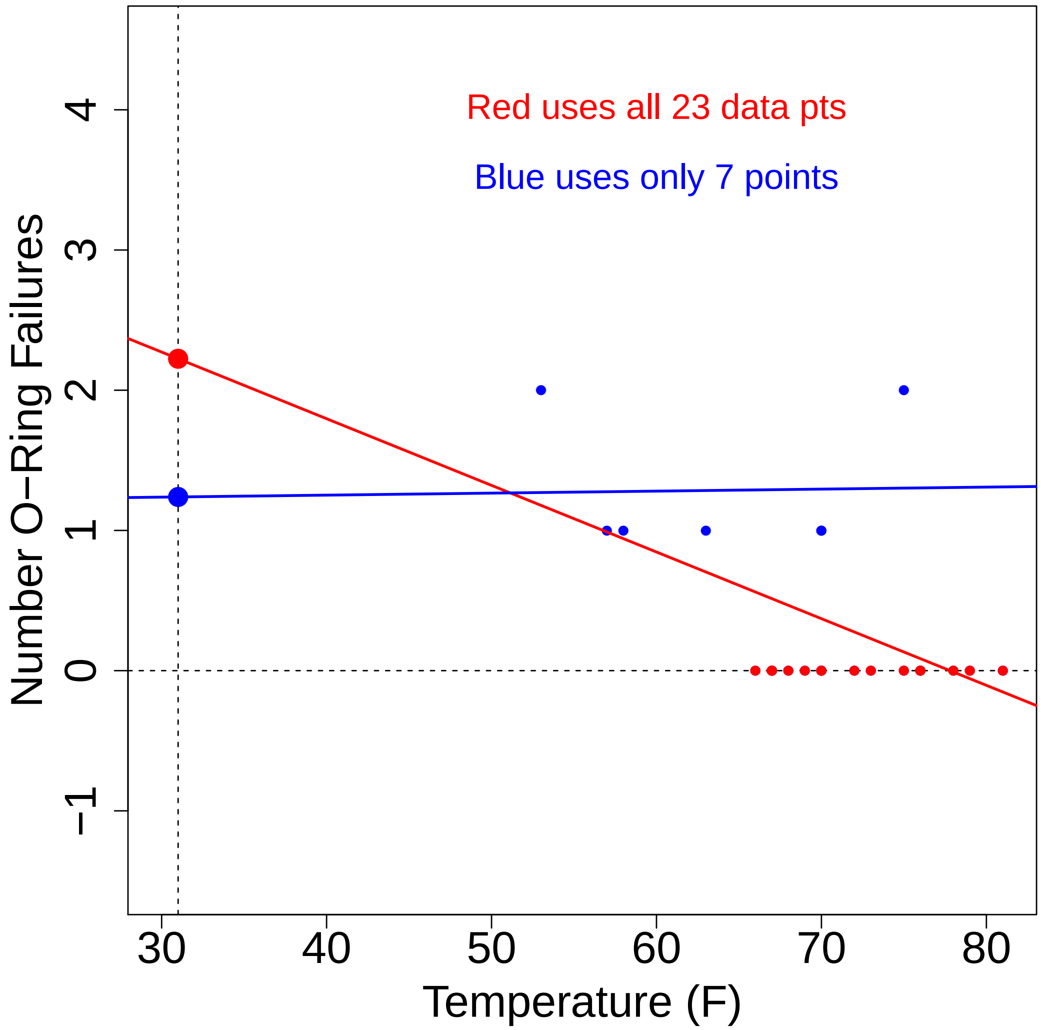
0 O-Ring Failures	(16):	66 67 67 67 68 69 70 70 72 73 75 76 76 78 79 81 (total 1154)
1 O-Ring Failures	(5):	57 58 63 70 70 (total 318)
2 O-Ring Failures	(2):	53 75 (total 128)

(a) What is the 95% confidence interval for Y at $x = 31$ degrees if you use all 23 data points? Add this to the plot below.

(b) What is the 95% confidence interval for Y at $x = 31$ degrees if you only use the 7 data points with at least 1 O-ring failure? Add this CI to the plot below.

(c) Does use of either dataset give you any confidence that a launch at 31 degrees might be safe? Why or why not?

Hint: Treat these data as though the temperature was not random, but "selected" by NASA.


Red uses all 23 data pts
Blue uses only 7 points
Number O−Ring Failures
Temperature (F)
4
3
2
1
0
−1
30
40
50
60
70
80