Chapter 4 – Applications of Statistical Inference

We'll end this course with this important ~100-page chapter.

Summary of Needed Theoretical Results (Section 4.1, pp. 156-159)

Theorem 4.1-1 states that if $X_1, X_2 \ldots X_n$ are n independent chi-square random variables with degrees of freedom $r_1, r_2 \ldots r_n$, then the sum $Y = X_1 + X_2 + \ldots + X_n$ has a chi-square distribution with degrees of freedom equal to $r_1 + r_2 + \ldots + r_n$.

Theorem 4.1-2 states that if $X_1, X_2 \ldots X_n$ are n independent Normal random variables with respective means $\mu_1, \mu_2 \ldots \mu_n$ and respective variances $\sigma_1^2, \sigma_2^2, \ldots \sigma_n^2$, then the random variable $Y = \Sigma^n a_k X_k$ has a Normal distribution with mean $\Sigma^n a_k\mu_k$ and variance $\Sigma^n a_k^2\sigma_k^2$.

Theorem 4.1-3 states that for a random sample of size n from a $N(\mu,\sigma^2)$ distribution with sample mean $\bar{X}$ and variance S^2, then

- **$\bar{X} \sim N(\mu, \sigma^2/n)$,**
- **$(n-1)S^2/\sigma^2 \sim \chi^2_{n-1}$, and**
- **$\bar{X}$ and S^2 are independent.**

Example 4.1-3 points out that whereas $\Sigma^n (X_k - \mu)^2 / \sigma^2 \sim \chi^2_n$, the random variable $(n-1)S^2/\sigma^2 = \Sigma^n (X_k - \bar{X})^2 / \sigma^2 \sim \chi^2_{n-1}$. One degree of freedom is lost in the latter expression due to estimation of μ by $\bar{X}$.

Theorem. If $Z \sim N(0,1)$, $U \sim \chi^2_r$, and Z and U are independent, then

$$T = \frac{Z}{\sqrt{U/r}}$$

has a Student's t distribution with r degrees of freedom. Areas/ probabilities for this distribution are given in <u>Table VI on p.334</u>.

Hence, for a random sample of size n, $\frac{\bar{X}-\mu}{S/\sqrt{n}}$ has a t_{n-1} distribution.

Theorem. Let $U_1 \sim \chi^2_r$, $U_2 \sim \chi^2_s$ and U_1 and U_2 are independent, then

$$F = \frac{U_1 / r}{U_2 / s}$$

has a (Fisher's or Snedecor's) F distribution with r and s degrees of freedom. Example 4.1.5 shows that for random Normal samples of size n and m, $F = \frac{S_X{}^2 \sigma_Y{}^2}{S_Y{}^2 \sigma_X{}^2}$ ~ F distribution with df = n-1 and m-1.

Some Confidence Intervals (Section 4.2, pp. 160-172)

In this Section, it is assumed that all samples are from Normal distributions, and are taken as approximations when this is not met.

A. Confidence intervals are given on p.160 for σ^2 and σ. The first is $[(n-1)S^2/b, (n-1)S^2/a]$ for $a = \chi^2_{(1-\alpha/2)}(n-1)$ and $b = \chi^2_{(\alpha/2)}(n-1)$, and the second interval just takes the square roots of the endpoints. To illustrate, for Example 4.2-1 on p.160, n = 13 (seeds), $12s^2 = 128.41$, for $\alpha = 10\%$, a = 5.226 and b = 21.03, so the 90% CI for σ^2 is [128.41/21.03, 128.41/5.226] = [6.11, 24.57] and the 90% CI for σ is [$\sqrt{6.11}$, $\sqrt{24.57}$] = [2.47, 4.96].

B. A confidence interval is given on p.161 for the ratio $\sigma_X{}^2/\sigma_Y{}^2$. It is

$$\left[\frac{1}{F_{\alpha/2}(n-1, m-1)} \frac{s_x{}^2}{s_y{}^2} \quad , \quad F_{\alpha/2}(m-1, n-1) \frac{s_x{}^2}{s_y{}^2} \right]$$

An illustration is provided in Example 4.2-2 on pp.161-2.

C. Confidence intervals for μ are addressed on pp.162-4. A two-sided 100(1-α)% confidence interval for μ is (top of p.163):

$$\left[\bar{x}-\frac{s\times t_{\alpha/2}(n-1)}{\sqrt{n}}\ ,\ \bar{x}+\frac{s\times t_{\alpha/2}(n-1)}{\sqrt{n}}\right]$$

Example 4.2-3 is related to the amount of butterfat for 20 cows, so for the 90% confidence interval for μ, the relevant t-statistic from p. 334 is t = 1.729, and the CI is [472.80, 542.20]. Our text neglects to provide the very important interpretation: we are 90% confident that the average butterfat of <u>all</u> such cows during this 305-day period is between 472.80 and 542.20 pounds. Example 4.2-4 performs a comparison of CI's for μ based on the above "t-method" and the "z-method" – the latter one assumes σ is known; both types of intervals have approximately 90% coverage. Whereas the above interval is two-sided, the text also provides one-sided intervals for μ at the top of p.165 when either a lower or upper bound is desired. For the lower bound case, this bound is $\bar{x}-s\times t_{\alpha}(n-1)/\sqrt{n}$ – note that the t-statistic is chosen here so that the α area is in one tail only.

D. Next, we want a CI for μ_X - μ_Y, the difference of means in this Normal setting. Recall the sample size for the X distribution is n and the sample size for the Y distribution is m. If we can assume that $\sigma_X^2 = \sigma_Y^2 = \sigma^2$, then the best estimator for σ^2 is the pooled estimator $s_P^2 = [(n-1)s_X^2+(m-1)s_Y^2]/[n+m-2]$. Then, for $t_0 = t_{\alpha/2}(n+m-2)$, the 100(1-$\alpha$)% CI for $\mu_X - \mu_Y$ is

$$\left[\bar{x}-\bar{y}-t_0 s_P\sqrt{\frac{1}{n}+\frac{1}{m}}\ ,\ \bar{x}-\bar{y}+t_0 s_P\sqrt{\frac{1}{n}+\frac{1}{m}}\right]$$

The derivation showing why T has a t-distribution is given on p.165. Example 4.2-5 assumes equal variances and finds the

95% CI for $\mu_X - \mu_Y$; here, df = 22 so t = 2.074, and s_P = 7.266 since $s_P^2 = [8 \times 60.76 + 14 \times 48.24] / 22 = 7.266^2$. Then, the 95% CI is [-3.65 , 9.05]. This means that we are 95% confident that μ_X exceeds μ_Y by as little as –3.65 and as much as 9.05; it's important to notice that this CI contains zero (more later).

E. In the previous paragraph, when we wish to set a CI for $\mu_X - \mu_Y$ but we cannot assume equal variances, then we use the statistic U given near the center of p.167 (and the CI derived there from), and the t-distribution with df = [v] with v given in the following formula on p.167; this is called the Welch method.

Confidence Intervals and Hypothesis Tests (Section 4.3, pp. 172-179)

Two of the most important uses of Statistics are setting CI's and also hypothesis testing (HT), and we now turn to HT. On p.172, the authors talk about wanting to assess a new method for teaching statistics based on concepts instead of formulae. In the past – the old method – say the average final grade score was 75 with an SD of σ = 10 points. We wish to test whether now – with the new method – average final test scores have *increased*. That is, we wish to test the null hypothesis H_0: μ = 75 vs. the alternative hypothesis H_A: μ > 75. We can do this by performing a (random) study, finding a test statistic (TS), and making a decision. Suppose we take a random sample of size n = 64. Based on our sample mean, if we wanted to set a one-sided lower-bounded 95% CI for μ, it would be $\bar{X}$ - 2.056; so if our sample mean was 77.47, the lower bound would be 75.41, and we would believe that the statistics reform method has indeed increased the mean (μ). More concisely, here the relevant test statistic (TS) is

$$Z = \frac{\bar{X} - \mu}{\sigma / \sqrt{n}}$$

For the H_A given above, we would reject H_0 and accept H_A if the TS $Z > z_\alpha$ from Table Vb on p.333 – that is our decision rule. If we do not reject H_0, then note that H_0 has not been 'proven' so we do not 'accept it', we simply 'fail to reject' it (or 'retain' it) – see the Remark on p.182. For the above example with $\alpha = 5\%$, $z_{0.05} = 1.645$, so the TS for $\bar{X} = 77.47$ is $Z = 1.976 > 1.645$, and we reject H_0 and accept that $\mu > 75$. Had we instead obtained a sample mean of $\bar{X} = 75.92$ then the TS would be $Z = 0.736 < 1.645$, and we would fail to reject H_0: $\mu = 75$.

There are two types of mistakes or errors we could make here. First, we could reject a true H_0 – this is called a type I error and its probability is α, called the significance level of the test. On the other hand, we could retain (fail to reject) a false H_0 – this is called a type II error and its probability is β. Calculating β depends upon the true mean since H_0 being false means H_A: $\mu > 75$ is true so we must specify the true μ. For example, to calculate β when $\mu = 76.5$, we obtain

$$\beta = P\left(\frac{\bar{X} - 75}{10/\sqrt{64}} < 1.645;\ \mu = 76.5\right)$$

$$= P\left(\frac{\bar{X} - 76.5}{10/\sqrt{64}} < 1.645 - \frac{76.5 - 75}{10/\sqrt{64}} = 0.445;\ \mu = 76.5\right) = 0.6718.$$

Since this type II error probability is quite high, the authors argue that we might want to design a more powerful study. How large a study is needed if we want $\alpha = 0.05$ and $\beta = 0.10$ (again at $\mu = 76.5$)?

We need to recalculate the critical region (CR) for $\bar{X}$: we have

$0.05 = P\left(\frac{\bar{X} - 75}{10/\sqrt{n}} > 1.645;\ \mu = 75\right) \rightarrow$ CR is $\bar{X} > 75 + 16.45/\sqrt{n}$. Also,

$0.10 = P(\bar{X} < 75 + 16.45/\sqrt{n};\ \mu = 76.5)$

$$= P\left(\frac{\bar{X}-76.5}{10/\sqrt{n}} < \frac{-1.5+16.45/\sqrt{n}}{10/\sqrt{n}};\ \mu = 76.5\right)$$

Hence, $\frac{-1.5+16.45/\sqrt{n}}{10/\sqrt{n}}$ = -1.28 → √n = 29.27/1.5 = 19.51. This yields n = 380.77, so we take n = 381 students. With this sample size, the significance level is α = 5% and the power is 1 – β = 90%.

In the center of p.176, the authors address a second illustration involving the breaking strength of steel bars (X). The old process yielded bars with an average breaking strength of μ = 50 and σ = 6, and researchers feel that the new process will yield bars with μ = 55 (and the same SD). Thus the null hypothesis here is H_0: μ = 50 and the alternative is H_A: μ = 55 – note here that since the alternative hypothesis contains only one value it is called a simple hypothesis. Our test statistic here is again based on $\bar{X}$, and the rejection or critical region is the set of sample points which produce large values of $\bar{X}$, such as for example $\bar{X} \geq 53$; the authors denote CR by C and the complement – the 'acceptance region' – by C'. Then, we can calculate α and β as on p.177, and this is also shown graphically on that page.

A decrease in the size of α leads to an increase in the size of β, and vice versa. Both α and β decrease if the sample size n is increased.

In both of the above examples, we considered processes that improve things, but often in scientific research, we only look for a change – and we are led to two-sided tests instead of the above one-sided tests. Returning to the above teaching example, we would continue to put the old test average into the null (H_0: μ = 75), but if we wanted to test for a change, the alternative would now be H_A: μ ≠ 75, and the critical region would need to divide the α in two (α/2 in the left tail and α/2 in the right tail). It's important to realize that this two-sided test is

exactly equivalent to the two-sided confidence interval considered in Section 4.2. The one- and two-sided tests here regarding μ are summarized in Table 4.3-1 on p.178. Also on p.178, the authors reiterate the connection between hypothesis testing (HT) and confidence intervals (CIs) for all of the cases considered: a single mean, two means, a single variance or SD, or two variances/SDs.

One-Parameter Basic Hypothesis Tests (Section 4.4, pp. 179-189)

In this section, we perform tests related to one parameter – either μ, σ^2, σ or p. First, (on p.179) we define the p-value.

> **The p-value associated with a test is the probability that we obtain a value of the test statistic that is at least as extreme (in the direction of the alternative) as the observed value of our test statistic; this probability is calculated assuming the null hypothesis is true.**

For example, in a Normal test of H_0: $\mu = 75$ versus H_A: $\mu > 75$ with $\sigma = 10$, n = 400, and $\bar{x} = 76$, the p-value is

$$\text{p-value} = P(\bar{x} > 76) = P\left(\frac{\bar{X}-\mu}{\sigma/\sqrt{n}} > \frac{76-75}{10/\sqrt{400}}\right) = P(Z > 2) = 0.0228$$

We hasten to point out that if the alternative was instead H_A: $\mu \neq 75$, then the p-value would be $2 \times P(Z > 2) = 2 \times 0.0228 = 0.0456$.

Note that our decision rule can be restated as follows:

- **If p-value < α, we reject H_0 and accept H_A**
- **If p-value $\geq \alpha$, we fail to reject H_0**

Example 4.4-1 on p.180 is concerned with the Z-test since it is assumed that σ is known and the hypotheses are related to the mean. Here, the null is H_0: $\mu = 60$ and the alternative is H_A: $\mu > 60$; $\sigma = 10$, n = 52, and $\bar{x} = 62.75$. Thus, the p-value is

$$\text{p-value} = P(\bar{x} > 62.75) = P\left(\frac{\bar{X}-\mu}{\sigma/\sqrt{n}} > \frac{62.75-60}{10/\sqrt{52}}\right) = P(Z > 1.98) = 0.0239$$

A graph of the p-value (the shaded right-hand tail probability) is given on p.181. Had we set for example $\alpha = 5\%$, we would reject H_0 here and conclude that $\mu > 60$. *Here, we need to assume normality but since the sample size is so large, this requirement is not too paramount.*

Example 4.4-2 on p.181 is concerned with the T-test since σ is not known (we use s to estimate σ) and the hypotheses are related to the mean. Here, the null hypothesis is H_0: $\mu = 4$mm and the alternative is H_A: $\mu \neq 4$mm; $\alpha = 0.10$, n = 9, $\bar{x} = 4.3$, s = 1.2. Thus, the p-value is

$$\text{p-value} = 2 \times P(\bar{x} > 4.3) = 2\times P\left(\frac{\bar{X}-\mu}{s/\sqrt{n}} > \frac{4.3-4}{1.2/\sqrt{9}}\right)$$

$$= 2 \times P(t_8 > 0.75) \approx 0.50 \text{ (exact value} = 0.4748).$$

The data do not suggest a significant departure from the hypothesized mean of 4mm. *We definitely need to assume normality here.*

Example 4.4-3 on p.183 is concerned with the T-test since these paired data are analyzed using the paired t-test. Let

W = Before time – After time

***We assume that these n = 24 differences come from a Normal distribution.* Here, $\alpha = 5\%$, the null hypothesis is H_0: $\mu_W = 0$ and the alternative is H_A: $\mu_W > 0$ – the alternative corresponds to the average 'Before value' exceeding the average 'After value'. Summarizing the**

data: n = 24, $\overline{w}$ = 0.079, s_W = 0.255; the test statistic is $\frac{0.079-0}{0.255/\sqrt{24}}$ = 1.518, so the p-value is $P(t_{23} > 1.518)$ – from Table VI, this value is between 0.05 and 0.10. The data do not suggest a significant difference between the Before and After average running times.

Example 4.4-4 on p.181 is concerned with the χ^2-test since the test is related to σ^2. The hypotheses are H_0: $\sigma^2 = 100$ and H_A: $\sigma^2 \neq 100$; $\alpha = 0.05$, n = 23, and $s^2 = 147.82$. The test statistic is $\frac{(n-1)S^2}{\sigma^2}$, which here is $\chi_{22}^2 = 32.52$, so the p-value is $2 \times P(\chi_{22}^2 > 32.52) \approx 0.18 > \alpha = 0.05$. These data do not suggest a significant departure from the hypothesized variance of 100. This result coincides with the sample variance (147.82) falling in the 95% CI for σ^2 given on p.184, viz, [88.42 , 296.18]. *We need to assume normality here.*

The final example is related to a Binomial proportion. The correct two-sided CI for p is given at the bottom of p.185:

$$\hat{p} \pm z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

On the other hand, to test the null hypothesis H_0: $p = p_0$, we use the test statistic:

$$Z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$$

Example 4.4-5 on p.186: p is the probability the tennis player is successful on her first serve after taking the lessons. The hypotheses are H_0: p = 0.40 (no change) and H_A: p > 0.40 (the lessons have yielded improvement). Here $\hat{p} = 92/200 = 0.46$, so $Z = \dfrac{0.46 - 0.40}{\sqrt{\dfrac{0.40 \times 0.60}{200}}} = 1.73$

is the test statistic. We get the p-value from the N(0,1) table – here, p-value = 0.0418 < 0.05 = α, so **we conclude that the lessons have significantly improved her first-serve success rate**.

Two-Parameter Basic Hypothesis Tests (Section 4.5, pp. 189-197)

Students may want to re-read Section 4.2 paragraphs B, D and E since CI's are related to HT's.

Here, we compare two means, two variances or two proportions from independently sampled groups using hypothesis testing. (Again, the null hypothesis will always contain the equal sign.) It is important to understand and appreciate the difference between:

(1) The situation considered in this section: two separate groups such as Male and Female students, etc., and
(2) The paired t-test situation in the last section, where the same person or unit is measured twice (e.g., before and after).

Example 4.5-1 on p.189 is related to the growth response of pea stems randomized to either lower hormone concentration [$X \sim N(\mu_X, \sigma^2)$ with sample size n = 11] or higher hormone concentration [sample size m = 13 and $Y \sim N(\mu_Y, \sigma^2)$]. Note that we are assuming Normality here and equal variances (more on the variances later). It is conjectured that higher hormone should result in higher average pea stem growth, so the hypotheses here are

$$H_0: \mu_X = \mu_Y \ (\mu_X - \mu_Y = 0)$$

$$H_A: \mu_X < \mu_Y \ (\mu_X - \mu_Y < 0)$$

The relevant test statistic (TS) is given on p. 189 in Equation (4.5-1). Here, the pooled variance estimate is

$$s_P^2 = [10 \times 0.24 + 12 \times 0.35]/22 = 0.5477^2,$$

so the TS is

$$t_{22} = \frac{\bar{x} - \bar{y}}{s_P\sqrt{\frac{1}{n} + \frac{1}{m}}} = \frac{1.03 - 1.66}{0.5477\sqrt{\frac{1}{11} + \frac{1}{13}}} = -2.8078$$

The p-value is $P(t_{22} < -2.8078) = P(t_{22} > 2.8078) \approx 0.005$. Since $p < \alpha = 0.05$, we reject H_0 and accept that the higher hormone concentration does appear to significantly increase average pea stem growth.

If we are interested in testing that two independent Normal variances are equal (H_0: $\sigma_X^2 = \sigma_Y^2$ or $\sigma_X^2/\sigma_Y^2 = 1$), we use either of the test statistics $F_{n-1,m-1} = s_X^2/s_Y^2$ or $F_{m-1,n-1} = s_Y^2/s_X^2$ – usually, it's easiest to put the larger sample variance over the smaller one.

In Example 4.5-3 on p.193, we return to the above pea growth example to test equality of the variances. Here, H_0: $\sigma_X^2 = \sigma_Y^2$ and H_A: $\sigma_X^2 \neq \sigma_Y^2$ and $\alpha = 5\%$. Our TS is $F_{12,10} = 0.35/0.24 = 1.458$. Since $1.458 < 3.62$, the TS is in the 'acceptance region' so we retain H_0. Alternatively, the p-value is $2 \times P(F_{12,10} > 1.458) > 0.10$, and since $p > \alpha$, we again fail to reject H_0. This helps give us confidence that in the equal-variance 2-sample t-test that we performed above, the assumption of equal variances could indeed be correct.

Finally, we turn to two independent-sample Binomial proportions, denoted p_1 and p_2. The respective sample sizes from the two groups are n_1 and n_2, and the random variables are Y_1 and Y_2. As in the previous section, let $\hat{p}_1 = Y_1/n_1$ and $\hat{p}_2 = Y_2/n_2$; for hypothesis tests with null H_0: $p_1 = p_2$ (= p), p is estimated by $\hat{p} = (Y_1+Y_2)/(n_1+n_2)$. Again, we have two slightly different procedures – one for CI's and one for HT's. Whereas CI's for (p_1 - p_2) use the standard error:

$$\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1}+\frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$$

HT's use the standard error:

$$\sqrt{\frac{\hat{p}(1-\hat{p})}{n_1}+\frac{\hat{p}(1-\hat{p})}{n_2}}=\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}$$

Example 4.5-4 on p.195 is related to manufacturing toggle levers during the day and night shift and the number of defects is counted. Here, $n_1 = n_2 = 1000$, the hypotheses are H_0: $p_1 = p_2$ and H_A: $p_1 < p_2$ (corresponding to a higher proportion of defects at night). Also, $\alpha = 0.05$. Since $\hat{p} = (37+53)/2000 = 0.045$, the estimated SE is

$$\sqrt{0.045\times 0.955\times\left(\frac{1}{1000}+\frac{1}{1000}\right)}=0.009271$$

Hence, the test statistic is Z = (0.037 – 0.053) / 0.009271 = -1.7258, so the p-value is P(Z < -1.73) = P(Z > 1.73) = 0.0418. Since $p < \alpha$, these data provide (marginal = not overwhelming) evidence that the proportion of defects during the night shift significantly exceeds the proportion of defects during the day shift.

Simple Linear Regression (Section 4.6, pp. 197-210)

It is often the case that in a study we measure two continuous variables (X and Y) on each person, and we wish to relate these variables. Here, we assume the relationship between X and Y is linear. Specifically, given the outcome $X = x_k$, we assume here that

$Y_k = \alpha_1 + \beta x_k + \varepsilon_k$; in this expression, x_k is the input or explanatory variable, Y_k is the response variable, α_1 is the true y-intercept, β is the true slope, and ε_k is called the 'error' – the deviation above or below the line. We'll actually find it more useful to assume the a slightly modified (but equivalent) version of this SLR model:

$$Y_k = \alpha + \beta(x_k - \bar{x}) + \varepsilon_k$$

Here, k = 1, 2 … n, and we'll assume that the x_k are fixed numbers (realizations of a random process), so only the Y_k are RV's since the ε_k are RV's. We assume the $\varepsilon_k \sim N(0, \sigma^2)$. This is a subtle, but important assumption: deviations above or below the line follow a Normal distribution and the variance does not depend upon the value of x; in *SLR modelling*, we try to validate this assumption if possible.

As a result of this distributional assumption, the authors show on p.199 that the MLE's for α and β are obtained by minimizing the objective function, $H(\alpha, \beta) = \Sigma \varepsilon_k^2 = \sum_{k=1}^{n}[y_k - \alpha - \beta(x_k - \bar{x})]^2$. Thus, the MLE's are $\boxed{\hat{\alpha} = \bar{y}}$ and $\boxed{\hat{\beta} = \frac{s_{xy}}{s_x^{\ 2}}}$ where (for all sums 1, 2 … to n)

$$(n-1)\,s_{xy} = \sum (x_k - \bar{x})(y_k - \bar{y}) = \sum x_k y_k - \frac{1}{n}(\sum x_k)(\sum y_k)$$ **, and**

$$(n-1)\,s_x^{\ 2} = \sum (x_k - \bar{x})^2 = \sum x_k^{\ 2} - \frac{1}{n}(\sum x_k)^2$$

A similar expression can easily be written for $(n\text{-}1)\, s_y^{\ 2}$ (used below). Since these MLE's also minimize the above sum of squares function $H(\alpha, \beta) = \Sigma \varepsilon_k^2$, economists also call them Least-Squares Estimators.

When we substitute the MLE's for α and β into the linear model, we obtain the fitted or predicted values:

$$\hat{y}_k = \hat{\alpha} + \hat{\beta}(x_k - \bar{x})$$

Then, the differences between the actual responses and the predicted or fitted responses, $e_k = y_k - \hat{y}_k = y_k - \hat{\alpha} - \hat{\beta}(x_k - \bar{x})$, are called the residuals; it is very important to understand/appreciate the difference between the (only theoretical) errors and the (observed) residuals.

When it comes to estimating the variance σ^2, the MLE of σ^2 (see p.200) is $\hat{\sigma}^2$ = SSResid / n, where the residual sum of squares, SSResid = Σe_k^2. Since this (MLE) estimator is biased, most software packages use the unbiased estimator, $\tilde{\sigma}^2$ = SSResid / (n - 2). We know that this latter estimator is unbiased since on p.205 (line 7), it is argued that SSResid / $\sigma^2 \sim \chi_{n-2}^2$, and thus E(SSResid / σ^2) = (n-2). Finally, an important descriptive measure is

$$r^2 = 1 - [\text{SSResid} / \{(n-1)*s_y^2\}]$$

This r^2 measure is the fraction of the total variation in the y's that is explained by the regression of y on x (see p.202). The (signed) square root of r^2 is r, the sample correlation coefficient on p.87: $r = s_{xy} / (s_x s_y)$. We sometimes use the relation $\hat{\beta} = r \times (s_y / s_x)$.

On pp. 203-4, the authors show that $E(\hat{\alpha}) = \alpha$ and $E(\hat{\beta}) = \beta$ – so that $\hat{\alpha}$ and $\hat{\beta}$ are unbiased estimators. Also, $Var(\hat{\alpha}) = \sigma^2 / n$, which is best estimated by $\tilde{\sigma}^2 / n$; $Var(\hat{\beta}) = \sigma^2 / \Sigma(x_k - \bar{x})^2$, and this variance is best estimated by $\tilde{\sigma}^2 / \Sigma(x_k - \bar{x})^2$. This latter expression appears under the square root sign in the denominator of T_1 in the center of p.205. If the variances are known, then the standardized parameter estimates have Normal distributions; but if we plug in the estimated variances, the standardized parameter estimates have t-distributions with df = n – 2.

Hypothesis testing and setting CI's is important in SLR modeling. If we accept that the slope is zero, then x is not a good linear predictor of y (and if the slope is non-zero then x is a good linear predictor).

Here is an example (from p.202 and the data are graphed on p.198). Here, x is the midterm grade, and y is the student's final grade.

x	y	x^2	xy	y^2	$\hat{y}$	$e = y - \hat{y}$	e^2
70	77	4900	5390	5929	82.56	-5.56	30.931016
74	94	5476	6956	8836	85.53	8.47	71.741645
72	88	5184	6336	7744	84.05	3.95	15.636006
68	80	4624	5440	6400	81.08	-1.08	1.160728
58	71	3364	4118	5041	73.66	-2.66	7.056424
54	76	2916	4104	5776	70.69	5.31	28.217302
82	88	6724	7216	7744	91.47	3.47	12.018265
64	80	4096	5120	6400	78.11	1.89	3.575957
80	90	6400	7200	8100	89.98	0.02	0.000305
61	69	3721	4209	4761	75.88	-6.88	47.371380
683	813	47405	56089	66731		≈ 0	217.709038

First, we are given the data – meaning the first two columns above. Next, we calculate columns 3-5. Then, we compute

$$\hat{\alpha} = 813 / 10 = 81.3$$

$$9\, s_{xy} = 56089 - 0.10\,(683)\,(813) = 561.1$$

$$9\, s_x^{\,2} = 47405 - 0.10\,(683)^2 = 756.1, \text{ and so}$$

$$\hat{\beta} = 561.1 / 756.1 = 0.742098.$$

Thus, the fitted line for these data is:

$$\hat{y}_k = 81.3 + 0.742098\,(x_k - \bar{x})$$

Plugging in each of the x values above gives the predicted values in the sixth column above, from which we find the residuals and squared residuals in the last two columns. Thus, here SSResid = 217.709, and the two estimates of σ^2 are $\hat{\sigma}^2 = 21.7709$ and $\tilde{\sigma}^2 = 27.2136 = 5.216668^2$.

Also, since $9s_y^{\,2}$ = 66731 – 0.10(813)^2 = 634.1, r^2 = 1 – 217.709/634.1 = 0.6567, and we say that 65.67% of the variability in the y's is explained by the regression of y = final grade on x = midterm grade (and 34.33% is not).

For testing purposes, we now need the estimated variances and SE's for our parameter estimates. For these data,

- **the estimate of the SE of $\hat{\alpha}$ is $\sqrt{\tilde{\sigma}^2/n}$ = 1.64966, and**
- **the estimate of the SE of $\hat{\beta}$ is $\sqrt{\tilde{\sigma}^2/\sum(x_k-\bar{x})^2}$ = 0.189716.**

To test H_0: $\alpha = 0$ versus H_1: $\alpha \neq 0$, the test statistic here is $t_8 = \frac{81.3-0}{1.64966}$ = 49.3, so the p-value = 2 × P(t_8 > 49.3) is near zero, and we conclude that the y-intercept is not zero (reject H_0 and accept H_1).

<u>More importantly</u>, to test for zero slope, H_0: $\beta = 0$ versus non-zero slope H_1: $\beta \neq 0$, the test statistic here is $t_8 = \frac{0.742098-0}{0.189716}$ = 3.912 and thus the p-value = 2 × P(t_8 > 3.912) = 0.0045 (via computer – using the t-table on p.334, we can only say that p-value < 2*(0.005) = 0.01 since 3.912 > 3.355). Therefore, even if α = 0.01 we reject H_0 and accept H_1 and we conclude that the slope is not zero: thus x = midterm grade is a good linear predictor of y = final grade. This is based on these data and contingent upon the above assumptions.

The final step in SLR modeling is to look at the residual plot (with the residuals on the y-axis and either x or the predicted values on the x-axis) as on p.203. We are looking for a *random pattern* and *check the constant variance assumption* by making sure that the top-to-bottom variability is the same as we move from left to right in the residual plot. These data look fine so the above conclusions seem justified.

Chi-Square Goodness of Fit Tests (Section 4.10, pp. 238-247)

A process results in one of k mutually exclusive and exhaustive events, $A_1, A_2, \dots A_k$, and let $p_s = P(A_s)$ for s = 1, 2 … k; thus, $\Sigma p_s = 1$. For n independent replicates from this process with observations $y_1, y_2, \dots y_k$ in the respective categories (so $\Sigma y_s = n$), the joint pmf of $Y_1, Y_2, \dots Y_{k-1}$ is the multinomial one:

$$f(y_1, y_2, \dots y_{k-1}) = \frac{n!}{y_1!y_2!\dots y_k!} p_1^{y_1} p_2^{y_2} \dots p_k^{y_k}$$

Here, $y_k = n - y_1 - \dots - y_{k-1}$. This distribution generalizes the Binomial.

Even though we don't use this distribution here, the GOF (goodness of fit) test is similarly based on k distinct and exhaustive categories; this test uses the test statistic (top of p.240):

$$Q_{k-1} = \sum_{s=1}^{k} \frac{(y_s - \xi_s)^2}{\xi_s}$$

In this expression, each $\xi_s = np_s$, and – provided each $\xi_s \geq 5$, this TS has a χ^2_{k-1} distribution when the null hypothesis is true. Also, H_0 contains the specified proportions, and the alternative is that at least one of the actual proportions differs from the specified value.

p.240, Ex.4.10-1 is concerned with a sequence of 51 digits and assessing whether the digits form a random pattern. We place into category 1 situations in which a given digit is followed by the same digit; into category 2 goes pairs where the following digit differs by one value (with 0 one unit away from 9), and all other situations fall into category 3. We can test for a random pattern in the digits by:

> **H_0: $p_1 = 1/10$, $p_2 = 2/10$, and $p_3 = 7/10$**
> **H_1: at least one of the p_s differs from those specified in H_0**

Notice how H_0 is derived by 'common sense' for this test!

Here, $y_1 = 0$, $y_2 = 8$ and $y_3 = 42$; $\xi_1 = 5$, $\xi_2 = 10$, and $\xi_3 = 35$; so the TS is

$$Q_2 = \frac{(0-5)^2}{5} + \frac{(8-10)^2}{10} + \frac{(42-35)^2}{35} = 5.0 + 0.4 + 1.4 = 6.8 > 5.991 = \chi^2_{0.05}(2).$$

Thus, for $\alpha = 5\%$, we reject the null hypothesis and conclude that these 51 digits do <u>not</u> appear to follow a random pattern.

Sometimes, we can combine the above GOF test with either a discrete or continuous distribution (as we did in the Application on pp.75-8). For example, in exercise 2.4-6 (p.79), we estimated the binomial parameters to be $\tilde{m} = 15$ and $\tilde{p} = 0.373118$ this yields the table:

	Categories						
	{0,1,2,3}	{4}	{5}	{6}	{7}	{≥ 8}	Total
y_s	10	7	13	13	10	9	62
ξ_s	8.0484	9.6372	12.619	12.518	9.5797	9.5999	62

The expected values above (ξ_s) were obtained using the above estimated m and p. Here, the TS is

$$Q_{6-1-2} = \frac{(10-8.0484)^2}{8.0484} + \frac{(7-9.6372)^2}{9.6372} + \frac{(13-12.619)^2}{12.619} + \frac{(13-12.518)^2}{12.518} + \frac{(10-9.5797)^2}{9.5797} + \frac{(9-9.5999)^2}{9.5999} = 1.2809 < 6.251$$

This TS has df $6-1-2=3$ since 2 parameters were estimated here; since the p-value > 0.10, we retain the Binomial(m=15, p=0.373118) distribution for these data. Another example (involving the Poisson distribution and one parameter to be estimated) is given on pp.243-4.

For a continuous illustration, we consider the Continuous Uniform distribution over the unit interval from 'a' to 'a+1', where 'a' is unknown. Our sample of n = 100 has a minimum sample value of 2.5, so we estimate 'a' to be 2.5. We then have the table:

	Sub-intervals				
	[2.50,2.75]	[2.75,3.00]	[3.00,3.25]	[3.25,3.50]	Total
y_s	20	32	31	17	100
ξ_s	25	25	25	25	100

Here, the TS is

$$Q_{4-1-1} = \frac{(20-25)^2}{25} + \frac{(32-25)^2}{25} + \frac{(31-25)^2}{25} + \frac{(17-25)^2}{25} = 6.96.$$

From Table IV on p.331, 0.025 < p-value < 0.05, so with $\alpha = 5\%$, we reject the translated continuous unit Uniform distribution here. See also the exponential example on pp.244-5.

Contingency Tables (Section 4.11, pp. 247-258)

In this section, we consider whether two or more multinomial distributions are equal (called tests of homogeneity), and then provide a test of independence of attributes; both situations use a chi-square and TS similar to the one used in the last section.

In Exercise 4.11-1 (p.256), a random sample of 300 Group A nurses and a random sample of 200 Group B nurses are categorized by the major type of work they do (six different categories):

	Category						
	1	2	3	4	5	6	Total
Group A	95 (88.8)	36 (37.2)	71 (68.4)	21 (23.4)	45 (46.2)	32 (36.0)	300
Group B	53 (59.2)	26 (24.8)	43 (45.6)	18 (15.6)	32 (30.8)	28 (24.0)	200
Total	148	62	114	39	77	60	500

Let p_{1A} be the population percentage of Group A nurses who work in Category 1 work and p_{1B} be the population percentage of Group B

nurses who work in Category 1 work. Similarly, define p_{1A} and p_{1B} for Category 2 work, and so on. Then, the relevant null hypothesis here is:

$$H_0: p_{1A} = p_{1B}, p_{2A} = p_{2B}, \ldots p_{6A} = p_{6B}$$

The alternative hypothesis is that at least one of these equalities is incorrect (i.e., that the percentage distributions are not identical for the two groups of nurses). Note that if the null hypothesis is true, then the best estimate of the proportion of category 1 nurses is 148/500 = 0.296, and the expected values (in green above) are obtained by multiplying 0.296 by 300 for the group A nurses and by 200 for the group B nurses. Similar calculations give all of the above expected values. Note that each of the expected counts exceeds 5.

The generic TS to use for this type of problem is:

$$q = \sum_{t=1}^{h} \sum_{s=1}^{k} \frac{(y_{st} - \xi_{st})^2}{\xi_{st}}$$

Under the Null, it has the χ^2 distribution with df = (h – 1) × (k – 1).

For this example, h = 2, k = 6, df = 5, and the TS is equal to:

$$\begin{aligned} q_5 = &\frac{(95-88.8)^2}{88.8} + \frac{(36-37.2)^2}{37.2} + \frac{(71-68.4)^2}{68.4} + \frac{(21-23.4)^2}{23.4} \\ &+ \frac{(45-46.2)^2}{46.2} + \frac{(32-36.0)^2}{36.0} + \frac{(53-59.2)^2}{59.2} + \frac{(26-24.8)^2}{24.8} \\ &+ \frac{(43-45.6)^2}{45.6} + \frac{(18-15.6)^2}{15.6} + \frac{(32-30.8)^2}{30.8} + \frac{(28-24.0)^2}{24.0} = 3.23 \end{aligned}$$

From Table IV, the p-value is between 0.10 and 0.90 (p = 0.665 from the computer). At the α = 5% level, the result is not significant. We retain the claim that the distribution of category percentages is the same for the two groups of nurses.

In the above example, we used the χ^2 test to test for commonality of two multinomial distributions (for the two groups of nurses); we can also use exactly the same testing methodology to see whether the row and column variables are independent (H_0) or associated (H_1).

In Exercise 4.11-5 (p.257), a random sample of 100 students are cross-classified by gender and the instrument they played:

	Instrument					
Gender	**Piano**	**Woodwind**	**Brass**	**String**	**Vocal**	**Total**
Male	4 (4.95)	11 (13.05)	15 (9.45)	6 (5.40)	9 (12.15)	45
Female	7 (6.05)	18 (15.95)	6 (11.55)	6 (6.60)	18 (14.85)	55
Total	11	29	21	12	27	100

The relevant null hypothesis here is:

H_0: Gender and Instrument of choice are independent
H_1: Gender and Instrument of choice are associated

Here, h = 2, k = 5, df = 4, and the TS is equal to:

$$q_4 = \frac{(4-4.95)^2}{4.95} + \frac{(11-13.05)^2}{13.05} + \frac{(15-9.45)^2}{9.45} + \frac{(6-5.40)^2}{5.40} + \frac{(9-12.15)^2}{12.15} + \frac{(7-6.05)^2}{6.05} + \frac{(18-15.95)^2}{15.95} + \frac{(6-11.55)^2}{11.55} + \frac{(6-6.60)^2}{6.60} + \frac{(18-14.85)^2}{14.85}$$

$$= 0.18 + 0.32 + 3.26 + 0.07 + 0.82 + 0.15 + 0.26 + 2.67 + 0.06 + 0.67 = 8.45$$

From Table IV, the p-value is between 0.05 and 0.10 (p = 0.076 from the computer). Although at the α = 5% level, the result is not significant, we point out that the results are marginally significant.

Also, the above calculation shows that the (marginal) deviation occurs among the Brass players.

We see that applications of the χ^2 test are far-reaching: yet another illustration related to continuous distributions is given on pp.250-1. It's important to note that this test is applicable when we are comparing three binomial distributions (p.254) or two multinomial distributions in a test of homogeneity (our first illustration above), on the one hand, and when we consider only one group but with two variables and do a test of independence, on the other hand. See the author's comments in the final paragraph on p.255.